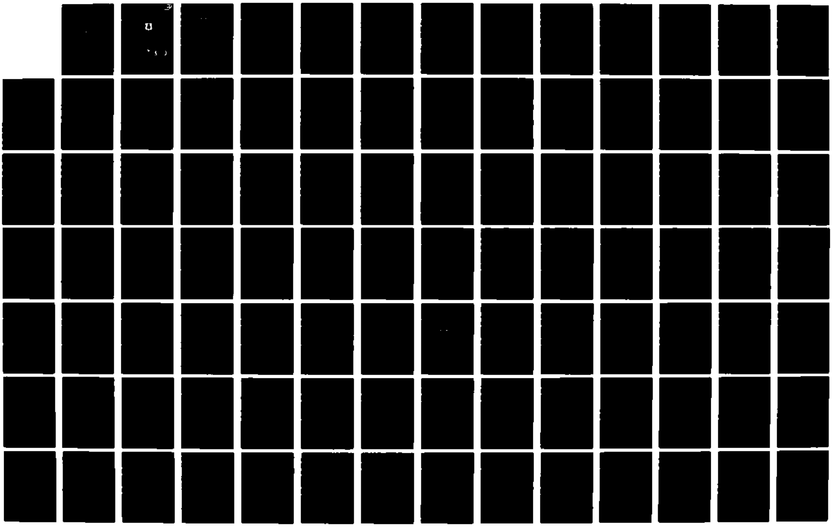
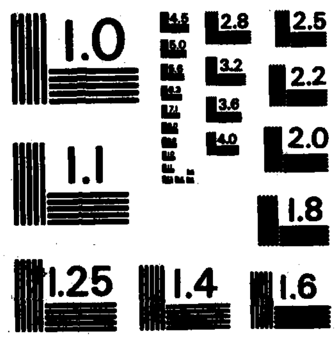
AD-A120 899 THEORY OF OPTIMUM RADIO RECEPTION METHODS IN RANDOM NOISE(U) FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OH L S GUTKIN 24 SEP 82 FTD-ID(RS)T-0784-82 1/7

UNCLASSIFIED F/G 9/4 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

FTD-ID(RS)T-0784-82

FOREIGN TECHNOLOGY DIVISION

THEORY OF OPTIMUM RADIO RECEPTION METHODS
IN RANDOM NOISE

by

L.S. Gutkin

DTIC
ELECTE
NOV 1 1982
S D
D

Approved for public release;
distribution unlimited.

AD A120899

DTIC FILE COPY

82 11 01 118

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

DTIC

FTD-ID(RS)T-0784-82

EDITED TRANSLATION

FTD-ID(RS)T-0784-82 24 September 1982

MICROFICHE NR: FTD-82-C-001255

THEORY OF OPTIMUM RADIO RECEPTION METHODS IN RANDOM NOISE

By: L.S. Gutkin

English pages: 637

Source: Teoriya Optimal'nykh Metodov Radiopriyema pri Fluktuatsionnykh Pomekhakh, "Sovetskoye Radio", Moscow, 1972, pp. 1-448

Country of origin: USSR
Translated by: SCITRAN
F33657-81-D-0263
Requester: USAMICOM
Approved for public release; distribution unlimited.

THIS TRANSLATION IS A RENDITION OF THE ORIGINAL FOREIGN TEXT WITHOUT ANY ANALYTICAL OR EDITORIAL COMMENT. STATEMENTS OR THEORIES ADVOCATED OR IMPLIED ARE THOSE OF THE SOURCE AND DO NOT NECESSARILY REFLECT THE POSITION OR OPINION OF THE FOREIGN TECHNOLOGY DIVISION.	PREPARED BY: TRANSLATION DIVISION FOREIGN TECHNOLOGY DIVISION WP-AFB, OHIO.

U. S. BOARD ON GEOGRAPHIC NAMES TRANSLITERATION SYSTEM

Block	Italic	Transliteration	Block	Italic	Transliteration
А а	*А а*	A, a	Р р	*Р р*	R, r
Б б	*Б б*	B, b	С с	*С с*	S, s
В в	*В в*	V, v	Т т	*Т т*	T, t
Г г	*Г г*	G, g	У у	*У у*	U, u
Д д	*Д д*	D, d	Ф ф	*Ф ф*	F, f
Е е	*Е е*	Ye, ye; E, e*	Х х	*Х х*	Kh, kh
Ж ж	*Ж ж*	Zh, zh	Ц ц	*Ц ц*	Ts, ts
З з	*З з*	Z, z	Ч ч	*Ч ч*	Ch, ch
И и	*И и*	I, i	Ш ш	*Ш ш*	Sh, sh
Й й	*Й й*	Y, y	Щ щ	*Щ щ*	Shch, shch
К к	*К к*	K, k	Ъ ъ	*Ъ ъ*	"
Л л	*Л л*	L, l	Ы ы	*Ы ы*	Y, y
М м	*М м*	M, m	Ь ь	*Ь ь*	'
Н н	*Н н*	N, n	Э э	*Э э*	E, e
О о	*О о*	O, o	Ю ю	*Ю ю*	Yu, yu
П п	*П п*	P, p	Я я	*Я я*	Ya, ya

*ye initially, after vowels, and after ъ, ь; e elsewhere. When written as ё in Russian, transliterate as yё or ё.

RUSSIAN AND ENGLISH TRIGONOMETRIC FUNCTIONS

Russian	English	Russian	English	Russian	English
sin	sin	sh	sinh	arc sh	$\sinh^{-1}$
cos	cos	ch	cosh	arc ch	$\cosh^{-1}$
tg	tan	th	tanh	arc th	$\tanh^{-1}$
ctg	cot	cth	coth	arc cth	$\coth^{-1}$
sec	sec	sch	sech	arc sch	sech^{-1}
cosec	csc	csch	csch	arc csch	csch^{-1}

Russian	English
rot	curl
lg	log

GRAPHICS DISCLAIMER

All figures, graphics, tables, equations, etc. merged into this translation were extracted from the best quality copy available.

TABLE OF CONTENTS

Part Four

Mathematical Methods of Statistics For Investigation of Optimum Signal Reception

/2

Gutkin, L. S. Teoriya optimal'nykh metodov radiopriyema pri fluktuatsionnykh pomekhakh ("Theory of Optimum Radio Reception Methods In Random Noise"). 2d Edition, Expanded and Revised, Moscow, "Sovetskoye radio," 1972, 448 pgs

The theory of optimum methods of reception of signals on the background of random noise, widely-used in development of any radioelectronic systems and devices based on reception and transmission of information (radar and radio-controlled, radio communications, radio telemetry, radio astronomy, television, and other systems), as well as electroacoustical and wire communications systems, is presented. Optimum linear and nonlinear filtration, binary and complex signal detection and discrimination, estimation of signal parameters, receiver synthesis for incomplete a priori data, special features of synthesis with respect to certain quality indicators, and other problems are examined.

The book is intended for a wide circle of specialists in radioelectronics and related branches of science and technology, and for senior students in higher educational institutions.

One table, 140 illustrations, 198 bibliographic entries.

FOREWORD

/7

The theory of optimum methods of reception of signals on a background of random noise (noise) is presented in this book. The term optimal also includes those methods of reception satisfying optimization previously formulated mathematically. This includes methods providing minimal mean square error in reproduction of the signal or of the messages the signal carries.

The foundations of this theory were laid in the seminal work of Academician V. A. Kotel'nikov Teoriya potentsial'noy pomekhoustoychivosti priyem pri fluktuatsionnykh pomekhakh ("Theory of Potential Reception Noise Immunity In Random Noise") published in 1946. The theory of optimal signal reception methods has developed continuously and rapidly since that time.

Since the first edition of this book was published (1961), the amount of literature devoted to the theory of optimal radio reception methods has increased by more than a factor of 5 and a multi-volume work would be required for a sufficiently-complete exposition of the contemporary state of this theory. Therefore, this book should be looked upon only as an introduction to the overall theory of optimal radio reception methods.

Compared to the first edition, this book includes four additional chapters (18, 20, 21, and 22) covering optimal nonlinear filtration, special features of the synthesis of optimal receivers when incomplete a priori information is available

(use of the theory of statistical solutions, game theory, the theory of nonparametric methods of statistics, and the theory of stochastic approximations), and situations where more than one quality indicator must be examined. Also, several minor changes have been made. In particular, about 10% of the material devoted to relatively-secondary problems has been eliminated to avoid a significant increase in the size of the book.

The following sequence is used, as was the case in the first edition, to simplify comprehension of the material and to approximate actual stages of development of this theory.

Some preliminary remarks required for understanding subsequent material are contained in Part I.

Part II is devoted to presentation of V. A. Kotel'nikov's theory of potential noise immunity and analysis of certain assumptions in this theory from the point of view of the present state of the problem. Kotel'nikov assumed that all signal parameters are precisely known at the point of reception (with the exception of the message subject to reproduction). Therefore, Part II in essence is devoted to optimal reception of precisely-known signals.

Optimal reception of signals comprising several unknown (random) parameters, /8 along with a useful message, is examined in Part III. The analytical method used here in essence is a direct offshoot of the Kotel'nikov method.

Part IV is devoted to analysis of optimal receivers and their properties using modern mathematical methods of statistics. Consequently, if Part III differs from Part II mainly by the type of signals examined, then Part IV differs from the other parts mainly by the analytical methods used. The methods presented in Part IV are general and moreover make it possible to use vast results obtained in mathematical statistics to find optimal receivers and analyze their properties.

As was the case in the first edition, the author strived wherever possible to simplify the presentation and make it accessible to a wide range of radio specialists and senior students in higher educational institutions.

The author expresses his sincere appreciation to A. Ye. Basharinov, I. A. Bol'shakov, V. A. Kordo, N. I. Belov, and Ya. Z. Tsypkin for much valuable advice taken into account in this book, as well as to Ye. M. Gutkina for great assistance in preparing the manuscript for publication.

PART ONE /9

PRELIMINARY INFORMATION

CHAPTER ONE. INTRODUCTION

1.1 Problem Formulation. Brief Historical Sketch

The problem of noise immunity, i. e., the problem of finding the best methods of radio signal reception in noise, is the fundamental and most complex radio reception problem. This is explained by the fact that the amount of noise increases simultaneously with receiver improvement, and signal reproduction requirements levied also increase.

Optimal reception methods are those which provide, in a certain sense, the best reception of signals or the messages they carry when noise is present. That which is included in the concept of "best" is referred to as optimization. Minimum mean square error, minimum complete error probability, and so forth may be applicable criteria.

In many instances, the purpose of the receiver is to reproduce only the message with which this signal is modulated, rather than the entire signal. Unless stipulated specially otherwise, the theory is constructed for modulated signals since unmodulated signals may be considered individual cases of modulated signals. Here, the basic

assumptions and results of the theory of optimal signal reception methods turn out to be valid for the most diverse signal types--modulated and unmodulated, radiotechnical, acoustical, wire communications signals, and so forth.

Therefore, the terms reception and receiver, rather than radio reception and radio receiver, will be used in future for commonality. However, the basic examples will be taken from radio reception since, in this instance, signals and noise usually have the most complex nature and protection against noise is the most pressing.

From a mathematical point of view, the task of finding the optimal receiver boils down in general terms to the following.

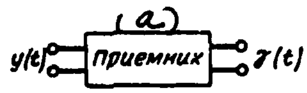


Figure 1.1. (a) -- Receiver.

Signal-plus-noise y(t) reaches receiver input (Figure 1.1):

$$y(t) = f[u_x(t),\ u_ш(t)], \tag{1.1}$$

where $u_x(t)$ -- signal carrying useful message x; $u_ш(t)$ -- interference ("noise").

For an unmodulated signal, $u_x(t) = \text{const} \cdot x(t)$. /10

In the simplest case, signal-plus-noise is simply their sum, i. e.

$$f[u_x(t),\ u_ш(t)] = u_x(t) + u_ш(t). \tag{1.2}$$

In this case, the noise is referred to as additive noise.

However, noise (prior to its arrival at the receiver) generally may distort the signal in a more complex manner, modulate one or several signal parameters (amplitude, phase, frequency, and so on), for example. Therefore, generally

expression (1.1) only denotes that oscillation y(t) arriving at receiver input is some function of signal-noise oscillations. We will use $\gamma(t)$ to denote an oscillation at a receiving device output.

We must find that receiver structure, i. e., that law of conversion of y(t) into $\gamma(t)$ in which $\gamma(t)$ will reproduce message x(t) the best (in a certain sense). Optimization must be formulated mathematically to solve this problem and some signal-plus-noise y(t) characteristics must be given, statistical characteristics, for instance.

The following basic problems are solved in the theory of optimum reception methods.

1. Selection of and foundation for the corresponding receiver optimizations.

2. Development of theoretical methods of finding the optimal receiver structure, i. e., receivers satisfying the selected optimization.

3. Investigation of the properties of optimal receivers, primarily, evaluation of their noise immunity, i. e., determination of the minimum signal to provide the given signal or message reproduction quality.

4. Comparison of optimal receivers determined theoretically with extant receivers (i. e., those already in use) to find the possibilities for and the advisibility of increasing the noise immunity of extant receivers.

5. Comparison of optimal receiver noise immunity for various known types of signals, in particular for various known methods of modulating these signals, to select the most noise-immune types of signals and methods of modulation.

6. Finding new, optimal signal types (modulation methods) which provide the greatest noise immunity as optimal receivers receive these signals.

Problems of finding and investigating the most noise-immune systems are examined not only in the theory of optimal reception methods, but in the general theory of communications as well. We will explain the link between these theories.

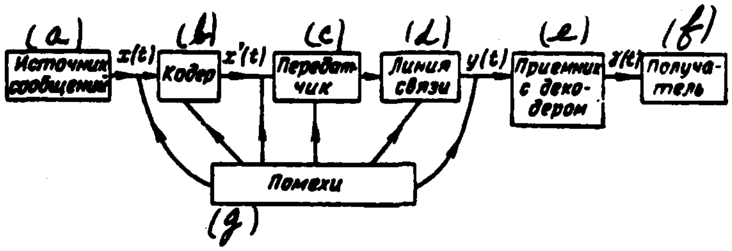


Figure 1.2. (a) -- Message source; (b) -- Coder; (c) -- Transmitter; (d) -- Communications line; (e) -- Receiver with decoder; (f) -- Recipient; (g) -- Noise.

A communications system (Figure 1.2) rather than an isolated receiver is examined in the general theory of communcations. Primary (useful) message x(t) is coded into secondary message x'(t), which modulates a transmitter. The /11 modulated signals arrive via a communications line (radio path, wire, and so on) to the receiver.

Without noise and other signal distortions, the oscillation at receiver input would have the form

$$y(t) = u_{x'}(t), \tag{1.3}$$

where x' = x'(t) -- distortionless coded message x(t).

Given ideal receiver operation, after demodulation and decoding, oscillation $\gamma(t)$ equalling or proportional to the primary message* would be isolated:

$$\gamma(t) = \text{const } x(t). \tag{1.4}$$

However, noise may be active throughout the entire communication system tract (Figure 1.2). It may cause distortion of secondary (encoded) message x'(t) and lead additionally to the fact that oscillation y(t) at receiver input will equal a complex function from $u_{x'}(t)$ and noise $u_{ш}(t)$, rather than $u_{x'}(t)$.

*For brevity, the assumption here is that the receiver simply is to reproduce the message.

Therefore, oscillation $\gamma(t)$ at receiver output differs from (1.4), i. e., the receiver will reproduce message x(t) with distortions.

The problem will be formulated in the following way in the general theory of communications: find those communications system unit (coder, transmitter, communications line, and receiver) construction principles providing the best (in a certain sense) message transmission from source to recipient when noise is present.

Consequently, in the general theory of communications, one seeks the optimum with respect to all permissible signal types (including all possible types of oscillation carriers and modulation and coding methods) and with respect to all possible methods of reception.

In the theory of optimal reception methods, one will seek the optimum primarily with respect to all possible methods of reception for the selected (usually sufficiently-broad) class of signals. Having obtained the corresponding results, it is then possible to find the optimum with respect to all possible types of oscillation carriers and modulation methods (within the range of the selected signal class), assuming here that the coding method is unchanged or generally assuming that there is no coding.

This problem was posed and successfully solved in several works, beginning with those of V. A. Kotel'nikov [1] and F. M. Woodward [2].

Using the theory of optimal radio reception methods, it is possible to take the final step--to find the optimum also with respect to message coding methods [182, 183]. Here, the problems that the theory of optimal reception methods solves in essence converge with the problems of the general theory of communications.

The general theory of communications evolved until the last decade mainly based on the classic theory of information. K. Shannon [185] developed the basic assumptions for this theory. Its main drawback is that it did not consider the importance of information for the problem posed. In particular, this hindered use of the classic theory of information for radar problems.

On the other hand, problems of optimal message coding and information transmission rates, being fundamental problems in the general theory of communications (transmission rate was examined only indirectly by limiting time T devoted to message reproduction) were not addressed in the theory of optimal radio reception methods until the last decade.

Therefore, until the last decade, the theory of optimal radio reception methods and the theory of information were two, virtually nonoverlapping, sections of the general theory of communications. However, in recent years, several works have appeared in which ideas lying at the foundation of both theories are addressed. Here, there has been a significant measure of success in surmounting limitations inherent in each of these theories, i. e., in considering the importance of information [133, 184, and others] and in using the theory of optimal reception methods for problems in message coding method optimization [133, 182, 183, 186, 187, and others]. Therefore, at present no sufficiently-distinct boundary yet exists between the theory of optimal reception methods and the theory of information and it is possible that, during the next decade, they will merge completely into a single theory of optimal message transmission and reception methods.

Now we will refine what we mean by the term receiver. If a receiver operates within an information transmission system (communications system), the Figure

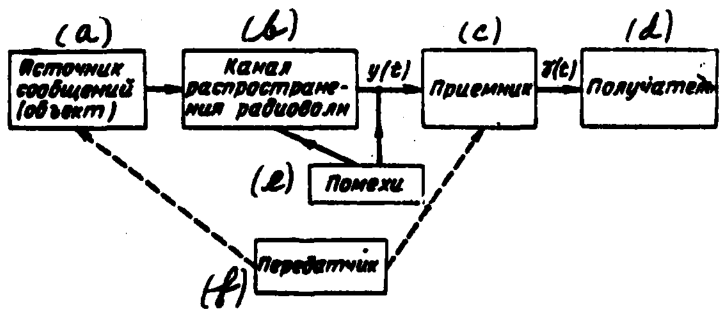


Figure 1.3. (a) -- Message source (object); (b) -- Radio wave propagation channel; (c) -- Receiver; (d) -- Recipient; (e) -- Noise; (f) -- Transmitter.

1.2 schematic describes its role. In the case of radar systems, the typical schematic is the one depicted in Figure 1.3. The message source is an object, whose parameters

(coordinates, speeds, and so on) must be determined. They are extracted by the receiver from the radio signal radiated or reflected by the object (by an aircraft, space vehicle, and so on). In several instances, a special radio transmitter painting the object is required to shape this radio signal.

The basic difference between a radar system and a communications system is that spontaneous modulations and coding of the signal carrier oscillation by the message is impossible in the former. Therefore, one may only seek the optimum with respect to possible signal carrier oscillation types and reception methods. In the case of passive radar, when an object's natural radiation is used to determine its parameters, optimization is possible only with respect to reception methods.

Any radio receiving device, regardless of the system for which it is /13
intended, will comprise three component parts (Figure 1.4): antenna-feeder device A, receiver, and output device (recipient). Signal-plus-noise $E(t, X, Y, Z)$ at antenna-feeder device input is a function not only of time, but of space coordinates (generally of three coordinates X, Y, and Z), and at this device's output, i. e., at receiver input it is a function of time $y(t)$.

A human or any automatic device (loudspeaker, oscillograph, electronic computer, and so forth) receiving reproduced message x may play the role of output device (recipient). The output device is not examined directly in the theory of optimal radio reception methods. It is considered only during mathematical formulation of requirements levied on receiver output effect $\gamma(t)$ (the requirement is for the mean square deviation $\gamma(t)$ from reproduced message $x(t)$ be minimal, for example).

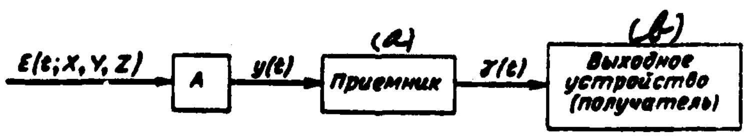


Figure 1.4. (a) -- Receiver; (b) -- Output device (recipient).

Either the antenna-feeder device and receiver combination or just the receiver is subjected to optimization during the synthesis process. In the latter case,

one will seek, as already noted above, the optimal receiver structure for given signal-plus-noise y(t) at its input.

It is evident that the best results are obtained if the antenna-feeder device and receiver combination, rather than the receiver, is optimized. This requires having statistical characteristics of random space-time process E(t, X, Y, Z) rather than of the random time process y(t). Here, both the mathematical formulation of the synthesis problem and its solution are complicated radically. /14

Complication of problem formulation means mathematical description of noise space-time characteristics. This description requires, in particular, very complete a priori information on the noise, which is lacking in many cases. However, in view of the great importance of the overall receiving device optimization process, including the antenna system as well, great attention has been devoted in recent years to space-time synthesis [134].

In future, where this is not stipulated, the discussion will involve only synthesis of the receiver, i. e., the term receiver input will be understood to mean process y(t) at antenna-feeder device output.

It should be noted that Figures 1.1--1.3 depict a single-channel system. The receiver in a multichannel system may have several inputs and (or) several outputs. Here, Figure 1.1, 1.2, and 1.3 will remain valid if scalar functions y(t), γ(t), x(t), and x'(t) are replaced by the corresponding vector functions

$$\left.\begin{aligned} \vec{y}(t) &= (y_1(t), \dots, y_m(t)), \\ \vec{\gamma}(t) &= (\gamma_1(t), \dots, \gamma_l(t)), \\ \vec{x}(t) &= (x_1(t), \dots, x_l(t)), \\ \vec{x}'(t) &= (x_1'(t), \dots, x_l'(t)), \end{aligned}\right\} \tag{1.5}$$

where m and l -- number of channels at receiver input and output, respectively.

Missions that modern radioelectronic systems accomplish often are so complex that the message reproduction process must be divided into two stages: primary signal-plus-noise processing and secondary processing. The division principle

may vary. For example, primary processing when information transmitted from an earth satellite is received may occur at various points on the earth's surface, while secondary processing occurs at a single computer coordinating center, which issues the final result $\gamma(t)$ of message $x(t)$ reproduction. Secondary processing in recent years mainly occurs by means of discrete (digital) equipment using specialized or standard digital computers.

If problem conditions require that the message reproduction system comprise two parts (primary and secondary processors), then the following variants for formulation of the problem for its synthesis are possible:

-- optimization of the primary processor for a given secondary processor;

-- optimization of the secondary processor for a given primary processor; /15

-- optimization of both processors simultaneously.

It is evident that highest-quality message reproduction may occur in the third variant. However, in several instances, this turns out to be too complex from a mathematical standpoint and from the point of view of technical realization. Therefore, primary and secondary processing system synthesis often occurs independently, i. e., in accordance with the first and second variants.

There is no difference from the point of view of principle among the aforementioned three variants since all may be reduced to the receiver synthesis problem depicted in Figure 1.1.

Here, in the first variant, the term $\gamma(t)$ (or $\overrightarrow{\gamma(t)}$) should be understood to mean effect $\gamma'(t)$ at primary processor output, i. e., to consider the secondary processor an output device.

In the second variant, the term $y(t)$ (or $\overrightarrow{y(t)}$) should be understood to mean signal-plus-noise at primary processor (or processors) input, rather than at antenna output.

In the third variant, the term receiver should be understood to mean the combination of primary and secondary processors.

In spite of the principled (ideological) commonality of the three variants, there are differences among them. Therefore, in recent years several special books [179, 180, and others] were devoted to the theory of optimal secondary processing.

It follows from what has been stated that the contemporary theory of optimal radio reception methods makes it possible to solve quite varied and complex problems arising during processing of essentially any radioelectronic and other information transmission and extraction systems. During the past 15 years alone, dozens of books [2-9, 11-13, 16, 17, 19, 29-32, 125, 139, 175-183, and others] have been devoted wholly or to a significant degree to this theory. If one sums up this material, eliminating repetition, he would end up with more than a dozen volumes. Therefore, this book should be considered only an introduction to the modern theory of optimal radio reception methods.

Creation of the theory of optimum linear filters, in which a linearity condition is imposed on a receiver system (filter) separating the signal from noise, historically preceded the theory of optimal reception methods.

The foundations for the theory of optimum linear filters were laid in the seminal works of Academician A. N. Kolmogorov [34] and N. Wiener [28] as early as 1941-1942. Wiener determined the requirements levied on the optimum linear filter transfer function stemming from the condition of obtaining the minimum mean square error at filter output.

Another criterion began to be employed after 1943 for synthesis of optimum linear filters, this being the criterion of the maximum ratio of peak signal voltage value to the mean square noise voltage value. The beginnings here were made /16
by the work of North [113], who solved this problem for white noise.

The works enumerated, in spite of their great value, touched upon optimal synthesis not of the receiver as a whole, but only of one of its elements--the linear filter. V. A. Kotel'nikov's work Teoriya potentsial'noy pomekhoustoychivosti ("Theory of Potential Noise Immunity"), published in 1946 as a doctoral dissertation, was the first work devoted to the problem of finding the optimal receiver and investigation of its properties. This work was published without major changes

in 1956 as a monograph [1]. In Kotel'nikov's work, almost all the basic problems in the contemporary theory of optimal reception methods were presented and solved to a certain degree for the first time. This work retains its fundamental character even today.

Kotel'nikov solved the problem for additive noise in the form of stationary gaussian noise (a stationary fluctuating oscillation with a normal law of distribution) and for the simplest optimizations (minimal mean square error and maximum message reproduction a posteriori probability criteria). In addition, he made several additional restrictions and assumptions during his analysis.

Some of his restrictions and assumptions successfully were removed in subsequent works. Cases of more complex noise were examined, particularly nonadditive noise leading to onset of parasitic random parameters in the received oscillation (i. e., of those parameters unknown at the point of reception not containing any information concerning the useful message received). Here, one primarily should note the work of F. Woodward [2 and 18], D. Middleton [16, 31, 118], and others. D. Middleton and others analyzed receiver optimality for generalized statistical criteria.

Mathematical methods of statistics--methods of testing statistical hypotheses and obtaining statistical estimates--began to be used widely after the early 1950's to find optimal receivers and to evaluate their properties.

The foundations of the contemporary theory of statistical estimates, i. e., estimation of distribution parameters of random values and random functions, had been laid as early as 1920-1930 by R. Fisher, and then developed in the works of J. Neyman, E. Pearson, and other mathematicians. Creation in 1930-1940 of the theory of testing statistical hypotheses is the work of Neyman and Pearson. These methods were developed significantly during World War II in works by A. Wald [29, 30], who developed the method of series analysis, the theory of statistical solutions, and game theory.

D. Middleton and several other authors in their works published after 1950 demonstrated that all these approaches may be used with great success in solution of problems of methods of optimal signal reception in noise.

In accordance with the aforementioned sequence in the development of the theory of optimal reception methods and for ease in understanding, the material /17 in this book is presented in the following manner.

Preliminary information required for construction of the theory of optimal signal reception methods (general nature of information, signals, and noise, optimum linear filter properties, and so forth) is presented in Part I.

A brief exposition of A. N. Kotel'nikov's theory of potential noise immunity is provided in Part II.

In Part III, the theory moves into the more general case of signals with parasitic unknown parameters. Here, as was the case in Part II, the theory will be constructed only for the simplest receiver optimizations--the criteria of maximum inverse probability and minimal mean square error.

Part IV is devoted to the search for optimal receivers with the aid of contemporary methods of mathematical statistics, which make it possible to solve the problem for broad classes of receiver optimization. It also is possible successfully to demonstrate with the aid of these approaches that the results obtained during analysis of optimal receivers intended for highly-reliable and -precise message reproduction essentially are identical for a broad class of optimizations, including the simplest criteria accepted by Kotel'nikov.

Therefore, in spite of the fact that the theory postulated in Parts II and III of the book are based on the simplest criteria, its results turn out to be valid for a broad class of optimizations if the requirements for message reproduction precision and (or) reliability are sufficiently high. However, problems which never arise in the simpler formulation of the tasks used in Parts II and III also are examined in Part IV. This category includes, in particular, impacts of the optimizations and type of a priori distributions on optimal receiver structure and properties, methods of receiver synthesis given incomplete a priori data, the special features with respect to several quality indicators, sequential signal detection, and others.

1.2 General Nature of Messages, Signals, and Noise

In general, useful signal $u_x(t)$ is a function of time modulated by message x. There are three characteristic cases, depending on message x type:

1. Reception of discrete messages.

2. Reception of individual values of continuous messages.

3. Reception of oscillations (filtration).

The <u>first case</u> occurs if the message may have only discrete values x_1, x_2, . . ., x_m. Here, the set of possible messages must be complete, i. e., include all possible message values. If problem conditions permit pauses, i. e., cases when there are no messages, then they must be considered cases in which zero /18 message x_o is transmitted. All messages are considered to be unknown magnitudes at the place of reception (if you assume that certain messages are known beforehand at the place of reception, then they are not considered messsages because they contain no information; reception of such "messages" in principle is unnecessary).

In the majority of cases, these unknown magnitudes during reception may be considered random magnitudes described by certain laws of distribution of a random variable.

In the simplest case of message reception, when its signal $u_k(t)$ corresponds to each message value x_k, while the results of reception of preceding messages are not considered during reproduction of message x_k in the receiver, the set of random messages (x_o, x_1, x_2, . . ., x_m) is characterized fully by the collection of unidimensional probabilities of these messages $P(x_o)$, $P(x_1)$, . . ., $P(x_m)$. More complex reception cases may require consideration also of composite probabilities $P(x_o, x_1)$, $P(x_1, x_2)$, and so on. All these distributions of a random message variable are referred to as a priori distributions since they are known (or assumed to be known) prior to the experiment, i. e., prior to investigation of signal-plus-noise y(t) arriving at receiver input. In future, the entire collection of a priori probabilities of message set (x_o, x_1, . . ., x_m) will be designated P(x).

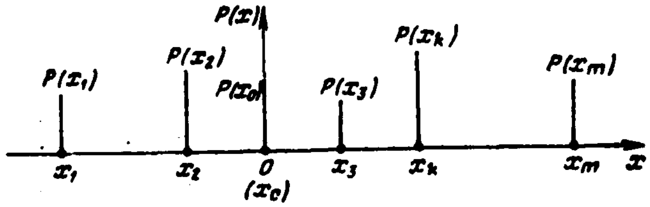


Figure 1.5

Distribution P(x) for a case where only unidimensional probabilities $P(x_k)$ are considered is depicted in Figure 1.5. Since the message set always is complete, then

$$\sum_{k=0}^{m} P(x_k) = 1. \tag{1.6a}$$

The <u>second case</u> occurs when the received message is a continuous magnitude having fixed value x during each experiment, i. e., during each observation cycle (T).

Actual messages usually are continuous functions of time x(t). However, if function x(t) changes slightly during observation cycle (T) (Figure 1.6), then one may consider that reception of individual values of continuous messages takes place. Here, message x may be looked upon during reception as an unknown magnitude,

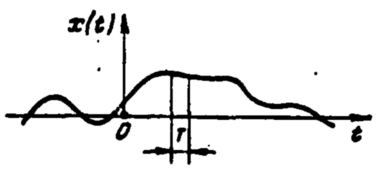


Figure 1.6

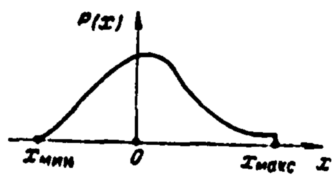


Figure 1.7

which may have any value within certain known limits-- from $x_{мин}$ to $x_{макс}$. In the majority of cases, this unknown magnitude during reception may be /19 considered a random magnitude having a certain a priori distribution. As usual, we will denote this distribution P(x) to obtain a series of standard messages

valid for any types of messages x, remembering however that there is a probability density rather than a probability in the case of continuous messages P(x).*

In a general case of reception, this distribution must include unidimensional and multidimensional probability densities. In the simplest cases, when the results of only one experiment (one observation cycle T) are considered as message x is being reproduced, it suffices to know unidimensional distribution P(x) similar to that depicted in Figure 1.7.

In this case

$$\int_{x_{\text{мин}}}^{x_{\text{макс}}} P(x)\,dx = 1. \qquad (1.6b)$$

The <u>third case</u> occurs if message x(t) changes over time so rapidly that it must be considered a time function even within the limits of one observation cycle (T). In this event, message x(t) reproduction during reception must be considered an unknown time function. In a majority of instances, this unknown function may be looked upon as a random function having a certain a priori distribution, which we again will designate P(x). But, P(x) now must be a multidimensional probability density characterizing the probability of certain realizations of function x(t).**

Thus, in all cases desired message x at the place of reception is an unknown magnitude or unknown time function. In a majority of cases, it is possible to consider this unknown magnitude (unknown function) as a random magnitude (random time function) characterized by some a priori distribution P(x). /20

P(x) is a combination of probabilities if message x is a discrete random magnitude. It is a probability density if message x is a continuous random magnitude.

*The possibility of obtaining a series of standard expressions for messages that are discrete random magnitudes, continuous random magnitudes, and continuous random time functions is stipulated by the fact that certain important relationships valid for probabilities (the law of probability multiplication, for example) remain valid for probability densities as well (unidimensional and multidimensional).

**That is, in this case P(x)dx is shorthand for the following expression: $P(x_1, x_2, \ldots, x_n)dx_1, dx_2, \ldots, dx_n$.

Distribution P(x) always is multidimensional (strictly speaking, infinitely-dimensional) during reception of oscillations. Distribution P(x) in the simplest case of reception may be unidimensional during reception of the individual values of continuous messages.

Now we will shift to the characteristics of the signals carrying message x. These signals may be divided into signals precisely known and signals with unknown parameters.

A signal is referred to as precisely known if message x is the only unknown parameter in oscillation $u_x(t)$, i. e., oscillation $u_x(t)$ is known precisely, given known x.

If, besides desired message x, a signal comprises some other unknown parameters--α_1, α_2, and so forth, not carrying any information concerning x, then it is referred to as a signal with unknown parameters and it is designated $u_{x,\alpha_1,\alpha_2}(t)$ or $u(x,\alpha_1,\alpha_2;\ t)$, respectively.

For example, let message x be transmitted with the aid of amplitude modulation, i. e.,

$$u_x(t) = u_0(1+x)\cos(\omega t+\varphi). \tag{1.7}$$

Such a signal, besides message x, is determined by three other parameters: u_0, ω, and ϕ. Therefore, signal (1.7) is considered precisely known if u_0, ω, and ϕ are known beforehand at the place of reception (i. e., prior to reception of combination y(t)).

If phase ϕ, along with x, also is known at the place of reception, then instead of (1.7) one must write:

$$u_{x,\varphi}(t) = u_0(1+x)\cos(\omega t+\varphi). \tag{1.8}$$

Signal parameters x, α_1, α_2, and so on during observation cycle (T) may

be both constant magnitudes and time functions. In many cases, unknown parasitic* parameters α_1, α_2, and so forth may be considered (or actually are) random magnitudes or random time functions. In these cases, signal $u(x, \alpha_1, \alpha_2, \ldots, \alpha_m; t)$ is referred to as a signal with random parameters.

Noise distorting a signal is divided into additive and nonadditive (modulating) noise. Additive noise, as noted in § 1.1, is the term used for noise $u_m(t)$, which is a component of combination y(t):

$$y(t) = u_x(t) + u_m(t).$$

All other noise is referred to as nonadditive noise.

Receiver internal noise is the typical and most important of the additive /21 noises.

Nonadditive noise, for example, includes noise arising in the process of radio wave propagation and reflection and causing signal parasitic modulation with respect to amplitude and phase. Nonadditive noise acts upon one or several signal parameters, i. e., causes signal parasitic modulation, so it may also be referred to as modulating noise.

Since nonadditive noise manifests itself in the change of certain parameters $\alpha_1, \alpha_2, \ldots, \alpha_m$, a signal distorted by such noise may be considered a signal with random parameters $\alpha_1, \alpha_2, \ldots, \alpha_m$. Therefore, signal-plus-noise y(t) may be written in the form

$$y(t) = u(x, \alpha_1, \alpha_2, \ldots, \alpha_m; t) + u_m(t) \quad (1.9)$$

when nonadditive noise is and is not present.

Parameters $\alpha_1, \alpha_2, \ldots, \alpha_m$ may be simply unknown when nonadditive noise is absent rather than being mandatory random magnitudes. One or several of these

*As noted in § 1.1, those signal parameters not carrying any information about message x that are unknown at the place of reception are called parasitic parameters.

parameters usually are random magnitudes or even random time functions when such noise is present.

As a rule, additive noise $u_m(t)$ is a random time function. Consequently, signal-plus-noise y(t) also must be considered a random time function.

Random time functions f(t), as is known, are characterized by multidimensional distributions of the type $P(f_1, f_2, \ldots, f_k, \ldots, f_n)$ where $f_1, f_2, \ldots, f_k, \ldots, f_n$ are values of function f(t) at individual moments in time $(t_1, t_2, \ldots, t_k, \ldots, t_n)$ of observation cycle (T), while magnitude n must be, strictly speaking, infinitely large (see § 1.3 on this point).

The random process is referred to as stationary (in the narrow sense of the word) if distributions $P(f_1, f_2, \ldots, f_n)$ do not change when the time reference changes (for any values of n). Otherwise, the process is not stationary.

Additive noise in many cases may be considered a stationary random process. In particular, the most important type of such noise--receiver internal noise, usually may be considered stationary. The majority of actual stationary processes have the property of ergodicity, i. e., averaging of process f(t) with respect to time given the same results as does statistical averaging for the given moment in time with respect to all possible realizations of the process.

We will designate averaging conditionally by means of the following symbols since we will in future be called upon often to use various types of averaging:

a) averaging with respect to time for infinitely-great time -- dashed line over the averaged magnitude:

$$\overline{f(t)} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t)\, dt; \tag{1.10a}$$

b) averaging with respect to time for finite time T -- enclosing the averaged magnitude in parentheses with subscript T

$$(f(t))_T = \frac{1}{T} \int_{0}^{T} f(t)\, dt; \tag{1.10b}$$

c) statistical averaging (finding the expected value) -- solid line above the average magnitude or the sign M (expected value) in front of this magnitude

$$\bar{f} = Mf.$$

If f -- continuous random magnitude, then

$$\bar{f} = Mf = \int_{A_f} fP(f)\,df, \tag{1.10c}$$

where A_f -- region of all possible values of magnitude f.

For ergodic processes

$$\bar{f}(\bar{t}) = \bar{f}. \tag{1.10d}$$

Such terminology as "white noise," "nonwhite noise," and "gaussian noise" are used in subsequent discussions to designate several of the most-important types of random noise. The term white noise is understood to mean a stationary fluctuating oscillation having an equidimensional frequency spectrum in the entire frequency range (from 0 to ∞).

If the noise spectrum is equidimensional in a restricted frequency range and equals zero outside this range, then such a fluctuating oscillation is referred to as noise with an equidimensional restricted spectrum. Noise with a non-equidimensional spectrum is referred to as non-white noise.

Since the white noise correlation function has the form of a delta function, such noise is referred to as uncorrelated noise. As opposed to this, non-white noise, including that with an equidimensional spectrum in a restricted band, is correlated noise.

Noise having a gaussian (normal) law of distribution (here, there may be any spectrum) is referred to as gaussian (normal) noise.

In accordance with this terminology, normal white noise is referred to as

a stationary random process having a normal law of distribution and an equidimensional frequency spectrum.

It should be noted that white noise, after passage through any real linear network, is converted at network output into a process with a normal law of distribution. This provides the basis to assume that the law of white noise distribution is normal.

1.3 Expansion of Time Functions Into Series With Respect to Orthogonal Functions

Expansion of time functions being investigated into series with respect to some known ("standard") orthogonal time functions is widely used in the theory /23 of optimal reception methods for mathematical description of signals and noise. Use of such expansions in many instances makes it possible significantly to simplify analysis and to make it clearer, especially if one is dealing with random time functions.

The most widespread is the Kotel'nikov expansion [33], which he obtained for time function f(t) having a spectrum restricted to frequencies from 0 to $f_{\text{в}}$. This expansion has the form

$$f(t)=\sum_{k=-\infty}^{\infty} f_k \psi_k(t), \tag{1.11a}$$

where $|k| = 0, 1, 2, 3, \ldots$;

$$\begin{gathered} f_k = f(k\,\Delta t); \quad \Delta t = \frac{1}{2f_{\text{в}}}; \\ \psi_k(t) = \frac{\sin 2\pi f_{\text{в}}(t-k\,\Delta t)}{2\pi f_{\text{в}}(t-k\,\Delta t)}. \end{gathered} \tag{1.11b}$$

Function $\psi_k(t)$ has the property of orthogonality, i. e.,

$$\int_{-\infty}^{\infty} \psi_k(t)\,\psi_l(t)\,dt = \begin{cases} 0 & \text{where } l \neq k; \\ \dfrac{1}{2f_{\text{в}}} & \text{where } l = k. \end{cases} \tag{1.11c}$$

In addition, it follows from (1.11a) and (1.11c) that

$$\int_{-\infty}^{\infty} f^2(t)\,dt = \frac{1}{2f_{\text{в}}}\sum_{k=-\infty}^{\infty} f_k^2. \tag{1.11d}$$

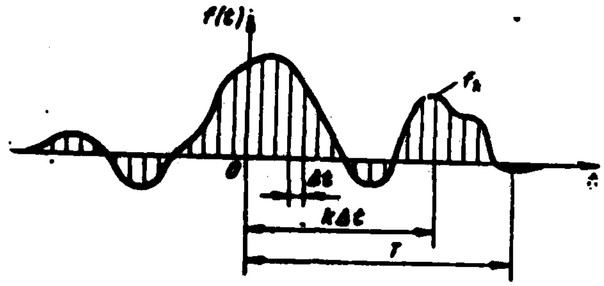


Figure 1.8

Coefficients f_k of series (1.11a) are values of function f(t) taken after discrete time intervals Δt (Figure 1.8).

If function f(t) with a restricted spectrum is considered only during finite time interval T (Figure 1.8), precise expansion (1.11) is replaced by the /24 following approximate expansion:

$$f(t) = \sum_{k=1}^{n} f_k \psi_k(t), \tag{1.12a}$$

where $f_k = f(k\Delta t)$; $\Delta t = \frac{1}{2f_{\text{в}}}$; $n = \frac{T}{\Delta t} = 2f_{\text{в}}T$;

$$\psi_k(t) = \frac{\sin 2\pi f_{\text{в}}(t - k\Delta t)}{2\pi f_{\text{в}}(t - k\Delta t)}. \tag{1.12b}$$

$$\int_0^T \psi_k(t)\,\psi_l(t)\,dt = \begin{cases} 0 & \text{where } l \neq k; \\ \dfrac{1}{2f_{\text{в}}} & \text{where } l = k. \end{cases} \tag{1.12c}$$

$$\int_0^T f^2(t)\,dt = \frac{1}{2f_B}\sum_{k=1}^{n} f_k^2. \qquad (1.12d)$$

Comparison of formulas (1.11) and (1.12) demonstrates that approximate relationships (1.12) are obtained from precise simple truncation of infinite sums, i. e., by restricting them by those values f_k of function f(t) which fall within the limits of time T (Figure 1.8). As will be demonstrated below, the expansion (1.12) error is slight if $n \gg 1$.

If function f(t) has relatively-narrow spectrum E(f) restricted by frequencies f_1 and f_2 (Figure 1.9), then it often is convenient to represent it in the form

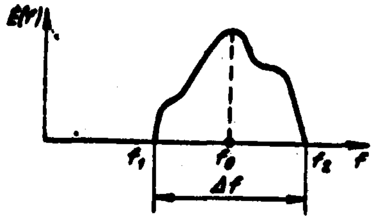


Figure 1.9

of an oscillation modulated with respect to amplitude and with respect to phase:

$$f(t) = r(t)\cos[2\pi f_0 t + \theta(t)], \qquad (1.13)$$

where $f_0 = \frac{f_1+f_2}{2}$ -- average frequency of the spectrum*, while r(t) and θ(t) -- amplitude (envelope) and initial phase of oscillation f(t), respectively.

Here, along with approximate expansion (1.12), it also is possible to use the following approximate expansion for time interval (T) [17, page 214]:

*It is possible to introduce f_0 by another method as well.

$$f(t)=\sum_{k=1}^{n_1}\psi_k'(t)\,r_k\cos(2\pi f_0 t+\theta_k), \tag{1.14}$$

where $r_k=r(k\Delta t)$; $\theta_k=\theta(k\Delta t)$; /25

$$\Delta t=\frac{1}{\Delta f};\quad n_1=\frac{T}{\Delta t}=\Delta f\cdot T;$$

$$\psi_k'(t)=\frac{\sin\pi\Delta f(t-k\Delta t)}{\pi\Delta f(t-k\Delta t)}.$$

Since $\psi_k'(t)$, similar to $\psi_k(t)$, are standard orthogonal functions, then function f(t) completely is determined by values r_k and θ_k of its amplitude and phase taken after time interval $\Delta t = 1/\Delta f$. The total number of these values is $2n_1 = 2\Delta fT$ (n_1 values of amplitude and n_1 values of phase), i. e., just as many as was the case for expansion (1.12), which uses instantaneous values of function f(t).

At the foundation of all these expansions is the assumption that the spectrum of the expanded function is restricted by band Δf, i. e., does not have components outside this band. This assumption naturally is an idealization since the spectrum of the oscillations at the output of actual physical systems always has several, albeit slight, "tails" rather than a complete cutoff outside any finite band.

It is possible to demonstrate (see [52] and [74], for example) that the assumption concerning the finitude of the spectrum width especially is undesirable from the point of view of principle when approximating random time functions since, strictly speaking, any random function must have an infinitely-wide spectrum. Therefore, to expand random time function f(t), it is possible to select interval width Δt between neighboring values $f(k\Delta t)$ of this function, stemming not from the width of the function f(t) spectrum, which should be considered infinitely large, but from its correlation interval τ_0, i. e., to assume

$$\Delta t=\tau_0. \tag{1.15}$$

Here, expansion into a series will be valid only in the event that this interval is much less than observation cycle T, i. e., only where

$$\frac{T}{\tau_0} \approx \frac{T}{\Delta t} \gg 1. \qquad (1.16)$$

In other words, it would be more natural when expanding random time functions to begin not with the assumption concerning the finitude of the width of their spectrum, but from the assumption concerning inequality with zero of their correlation interval τ_0.

However, expansion methods based on the assumption concerning the spectrum constraint are used more widely both for determinate and for random time functions. A large number of results important for practice and proven have been obtained with their help. Therefore, in future, we will use these very expansions to present the material.

We will examine the simplest of them in more detail, specifically Kotel'nikov expansions (1.11) and (1.12).

As already noted, expansion (1.11) is precise if function f(t) has a spectrum restricted by band $f_в$. This formula provides a relative error, whose mean square has a magnitude on the order of ΔE/E [66], for real functions with an unrestricted band. Here, E -- complete energy of the spectrum (unrestricted with respect /26 to width), while ΔE -- energy of the spectrum "tail", i. e., that part of it located outside band $f_в$. Consequently, the expansion (1.11) error is infinitesimally small if only a relatively-small amount of the studied function's spectrum energy is concentrated outside band $f_в$.

Expansion is used in the theory of optimal reception methods mainly for oscillation y(t) arriving at receiver input. Here, usually it is possible to consider function y(t) spectrum width $f_в$ as large as desired, and formula (1.11) here may be considered very precise. In addition, in a majority of instances, use of expansions (1.11) and (1.12) is only an intermediate mathematical drill--during the subsequent computations following expansion, an opposite transition occurs from series (sums) to convolute expressions (integrals), in which spectrum width is not included, and therefore can be a large as desired.

Thus, in a majority of cases, use of expansion (1.11) will not introduce any errors, major or otherwise, in spite of the fact that real functions have a theoretically-unrestricted spectrum. Expansion (1.12), as opposed to (1.11), is approximate, even if function f(t) has a restricted spectrum.

Actually, formula (1.12) was obtained from precise (given a restricted band) formula (1.11) by means of simple truncation, i. e., by discarding those parts of function f(t) which are outside interval (T) (Figure 1.8). Therefore, expansion (1.12) will not provide an error compared to (1.11) only when entire function f(t) actually is concentrated in interval (T). But the function, concentrated in a finite time interval, in principle may not have a restricted spectrum. Thus, expression (1.12) in principle may not be precise if one simultaneously assumes that both spectrum width f_B and observation cycle T duration are finite, i. e., if the number $2f_BT$ of discrete values of this function are finite.

For the formula (1.12) error to be slight, this condition must be met

$$n=2f_BT\gg 1. \tag{1.17}$$

This inequality essentially may be satisfied by increasing both T and f_B.

Formula (1.12) coincides with (1.11) when $T\to\infty$ (and with finite f_B) and only the aforementioned shortcomings linked with the assumption of the finitude of spectrum width remain.

If $f_B\to\infty$, then, as is easy to demonstrate, expansion (1.12) will become precise even for a finite T. Actually, it follows from expressions (1.12a) and (1.12b) that, at discrete moments kΔt, equality

$$f(t)=\sum_{k=1}^{n} f_k\psi_k(t)$$

is satisfied precisely since /27

$$\frac{\sin 2\pi f_B(t-k\Delta t)}{2\pi f_B(t-k\Delta t)}=\begin{cases}1 & \text{where } t=k\Delta t;\\ 0 & \text{where } t=l\Delta t,\ (l\neq k).\end{cases}$$

Consequently, the inexactitude of series (1.12a) may manifest itself only <u>within</u> each interval Δt, and not at its ends. But, when $f_B\to\infty$, each interval Δt will strive towards zero (Figure 1.8), the edges of this interval merge, and,

consequently, series (1.12a) will become precise not only for discrete, but also for all moments in time in the ranges examined from 0 to T.

Thus, when $f_s \rightarrow \infty$, expansion (1.12) is made precise even during finite observation cycle T.

As already noted above, in the theory of optimal reception methods, expansion of function f(t) into series (1.12) usually is only that intermediate mathematical operation in which the spectrum f_s width is not included and, therefore, it may be considered as large as desired. Here, use of expansion (1.12) provides no errors of any kind. Its use is convenient because function f(t) completely is determined by finite number n of its discrete values f_k, usually referred to as sample values. These values f_k also may be considered the coordinates of function f(t) in n-dimensional space (see § 6.3).

Use of expansion (1.12) is especially useful if f(t) is a random time function. Actually, in accordance with this expansion, function f(t) completely is determined

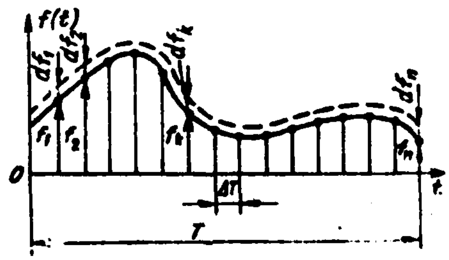


Figure 1.10

by its n ordinates f_1, f_2, . . ., f_n taken at discrete moments in time with some interval $\Delta t = 1/2f_s$ (Figure 1.10). Therefore, probability dP, such that realization of $f_1(t)$ of random time function f(t) will be fixed within the boundaries of an infinitely-thin layer restricted in Figure 1.10 by dotted and solid lines, equals the composite probability, such that ordinates f_1, f_2, . . ., f_n will be located accordingly within the limits

$$f_1 \div f_1 + df_1,\ f_2 \div f_2 + df_2, \ldots, f_n \div f_n + df_n,$$

i. e.,

$$dP = P(f_1, f_2, \ldots, f_n)\, df_1\, df_2 \ldots df_n, \tag{1.18}$$

where $P(f_1, f_2, \ldots, f_n)$ -- n-dimensional probability density.

Consequently, the random time function completely is characterized by n-dimensional probability density $P(f_1, f_2, \ldots, f_n)$ given for the entire range of possible values of ordinates $f_1, f_2, \ldots, f_n$.

We will find as our example the n-dimensional distribution for normal white noise, i. e., for stationary random time function $u_m(t)$, having a normal distribution and equidimensional frequency spectrum. It should be assumed that the spectrum /28 of this noise is limited by the band $(0 \div f_s)$ in order for expansion (1.12) to be applicable.

The following relationships are valid for such noise:

$$\overline{\overline{u_m}\,(t)} = \overline{u}_m = 0; \tag{1.19a}$$

$$\overline{\overline{u_m^2}\,(t)} = \overline{u_m^2} = N; \tag{1.19b}$$

$$\left.\begin{array}{l} S_m(f) = N_0 = \dfrac{N}{f_s} \text{ where } \quad f = 0 \div f_s; \\ S_m(f) = 0 \text{ outside these limits;} \end{array}\right\} \tag{1.19c}$$

$$W_1(u_{mk}) = \frac{1}{\sqrt{2\pi N}}\, e^{-u_{mk}^2/2N}, \tag{1.19d}$$

where $S_m(f)$ -- noise energy spectrum; N -- noise voltage mean square equalling its dispersion (since $\overline{u}_m = 0$); parameter N may be considered just like the noise power density, i. e., power developed by voltage $u_m(t)$ across a 1 ohm resistance; N_0 -- noise power density arriving per unit of band; $W_1(u_{mk})$ -- unidimensional law of noise distribution.

We will find the noise normalized correlation function $\rho(\tau)$ where

$$\rho(\tau)=\frac{\overline{u_m(t)\,u_m(t+\tau)}}{\overline{u_m^2(t)}}. \tag{1.20}$$

Function $\rho(\tau)$ is coupled with energy spectrum $S_m(f)$ by the following known relationship [20]:

$$\rho(\tau)=\frac{1}{\overline{u_m^2(t)}}\int_0^\infty S_m(f)\cos 2\pi f\tau df. \tag{1.21}$$

Considering relationship (1.19c), we obtain

$$\rho(\tau)=\frac{1}{N}\int_0^{f_s} N_0\cos 2\pi f\tau df=\frac{\sin 2\pi f_s\tau}{2\pi f_s\tau}. \tag{1.22}$$

A curve with respect to this formula is plotted in Figure 1.11, from which it is evident that there is no cross-correlation of voltage values $u_m(t)$ separated by interval $1/2f_s$. Hence it follows that, when noise voltage is represented as series (1.12), ordinates $f_1, \ldots, f_n$ included in this series should be considered uncorrelated. In addition, as follows from (1.19d), these ordinates are subordinate to the normal law of distribution. But, it is known from the theory of probabilities that, given a normal law of distribution, the absence of correlation signifies also an absence of statistical conjunction, i. e., statistical independence [20].

Therefore, in a case of normal white noise with a limited band, ordinate values $f_1, f_2, \ldots, f_n$ are statistically independent magnitudes and n-dimensional probability density $P(f_1, \ldots, f_n)$ equals the product of the equidimensional probability densities:

$$P(f_1, f_2, \ldots, f_n)=P(f_1)\,P(f_2)\ldots P(f_n). \tag{1.23}$$

Therefore, in accordance with (1.19d) and (1.23), it is possible to write

$$W_m(u_m)=P(u_{m1}, u_{m2}, \ldots, u_{mn})=$$
$$=\frac{1}{(\sqrt{2\pi N})^n}\exp\left(-\frac{1}{2N}\sum_{k=1}^{n} u_{mk}^2\right). \tag{1.24}$$

where $n = 2f_n T$.

Considering (1.12d), it is possible to represent expression (1.24) also in the following form:

$$W_m(u_m) = \frac{1}{(\sqrt{2\pi N})^n} \exp\left(-\frac{1}{N_0}\int_0^T u_m^2(t)\,dt\right). \tag{1.25}$$

It follows from expressions (1.24) or (1.25), in particular, that the greatest probability density corresponds to zero realization of noise voltage $u_m(t)$, i. e., to that realization which equals zero in entire observation cycle T.

CHAPTER TWO /29

OPTIMUM LINEAR FILTERS

2.1 Introductory Notes

The problem of optimum linear filters is illuminated in detail in several domestic and translated books [3, 8, 10-12, 136, and others]. Therefore, we will dwell briefly only on two of the most-characteristic variants of all those ways of formulating the problem of optimum linear filters--a filter creating the maximum signal-to-noise ratio at its own output and a filter providing minimal mean square error in signal reproduction.

As will be demonstrated in § 2.5, the basic field of use for the first type of filter (i. e., those creating the maximum signal-to-noise ratio) is filtration of modulated signals, i. e., filtration preceding signal detection (demodulation). Here, the job of the optimum filter is to provide best separation, not of /30 entire modulated signal $u_x(t)$ on the noise background, but only useful message x carried on it, which is a parameter of this signal.

It will be demonstrated in § 2.5 that the first type of optimum filter in many cases is a component part of the optimum receiver.

It is advisible to use filters of the second type, as shown in § 2.5, in

those instances when entire signal $u_c(t)$, rather than one (or several) of its parameters, is the useful message subject to separation from the noise. Therefore, such filters may be used, for example, in receiver stages connected beyond the detector (demodulator), as well as in various automated and telemechanical devices.

2.2 Linear Filters Providing Minimum Mean Square Error

Classic formulation of this problem, posed by N. Wiener as early as 1941 [28],

Figure 2.1. (a) -- Line filter.

involves the following (Figure 2.1). The sum of signal and noise arrives at stationary linear system (filter) input:

$$y(t) = u_c(t) + u_{ш}(t). \tag{2.1}$$

Both noise and signal are considered stationary random processes with zero mean values and known correlation functions $R_c(\tau)$ and $R_ш(\tau)$. If oscillations $u_c(t)$ and $u_ш(t)$ are cross-correlated, then the function of their cross correlation $R_{cш}(\tau)$ also is assumed to be known and does not depend on the time reference.

Since a filter is a stationary linear system according to the condition, then it is determined completely by its transfer function $K(j\omega)$ or pulse unit step function $\eta(t)$. As is known, $K(j\omega)$ and $\eta(t)$ are linked by a Fourier transform:

$$\eta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} K(j\omega) e^{j\omega t} d\omega, \tag{2.2}$$

$$K(j\omega) = \int_{-\infty}^{\infty} \eta(t) e^{-j\omega t} dt. \tag{2.3}$$

For a physically-realized filter, this condition must be satisfied* /31

$$\eta(t)=0 \quad \text{where} \quad t<0, \tag{2.4}$$

and relationship (2.3) takes the form

$$K(j\omega)=\int_0^\infty \eta(t)\, e^{-j\omega t}\, dt. \tag{2.3a}$$

The task of the linear filter, along with removing noise from the signal, also may include some sort of linear conversion, such as amplification, differentiation, integration, and so on. Therefore, generally speaking, certain time function h(t) coupled with the signal by a given linear conversion must be obtained at filter output when noise is absent. For instance:

$h(t)=\frac{du_c(t)}{dt}$, if differentiation is required;

$h(t)=\int_0^t u_c(t)\,dt$, if integration is required;

$h(t)=u_c(t-\Delta t)$, if a time shift by Δt is required;

$h(t)=au_c(t)$, if amplification by a factor of a is required;

$h(t)=u_c(t)$, if simple signal reproduction is required. Consequently, in the general case, there is a requirement to obtain

$$h(t)=\mathcal{D}(p)\,u_c(t), \tag{2.5}$$

where $\mathcal{D}(p)$ -- linear transformation; p = d/dt.

Oscillation $\gamma(t)$ at linear system output is linked with oscillation y(t) at input by known relationship

$$\gamma(t)=\int_{-\infty}^{\infty} y(t-\tau)\,\eta(\tau)\,d\tau. \tag{2.6}$$

*In addition, function $\eta(t)$ must (at the corresponding rate) strive towards zero when $t\to\infty$; but, in a majority of instances, condition (2.4) is limiting.

In the case of a physically-realized filter, this relationship, in accordance with (2.4), takes the form

$$\gamma(t)=\int_0^\infty y(t-\tau)\,\eta(\tau)\,d\tau. \tag{2.6a}$$

Therefore, when both signal and noise are present at input, at output there is

$$\gamma(t)=\int_0^\infty [u_c(t-\tau)+u_m(t-\tau)]\,\eta(\tau)\,d\tau. \tag{2.7}$$

The required oscillation at filter output is h(t). Therefore, error $\varepsilon(t)$, obtained at filter output equals /32

$$\varepsilon(t)=\gamma(t)-h(t). \tag{2.8}$$

The mean square of this error equals

$$\overline{\varepsilon^2}=\overline{[\gamma(t)-h(t)]^2}, \tag{2.9}$$

where formula (2.7) determines $\gamma(t)$.

The optimal filter is one with minimum magnitude $\overline{\varepsilon^2}$. Consequently, mathematically the problem boils down to finding that type of pulse transient characteristic $\eta(t)$ included in formula (2.7) in which magnitude (2.9) is minimum. Wiener solved this problem using calculus of variables approaches; it was found that desired characteristic $\eta(t)$ of the optimim linear filter must be solution of the following integral equation (see [3], for example):

$$\int_0^\infty R_y(\tau-t)\,\eta(t)\,dt=R_{yh}(\tau)\ \text{where}\ \ \tau\geqslant 0, \tag{2.10}$$

$$\left.\begin{aligned} \text{where}\quad & R_y(\tau)=\overline{y(t)\,y(t+\tau)} \\ \text{and}\quad & R_{yh}(\tau)=\overline{y(t)h(t+\tau)} \end{aligned}\right\} \tag{2.11}$$

are correlated functions assumed to be known.

If the signal and noise statistically are independent, then

$$\left.\begin{aligned} R_{yh}(\tau)&=\overline{u_c(t)\,h(t+\tau)}, \\ R_y(\tau)&=R_c(\tau)+R_m(\tau), \end{aligned}\right\} \tag{2.12}$$

where

$$R_c(\tau)=\overline{u_c(t)\,u_c(t+\tau)}$$

and

$$R_m(\tau)=\overline{u_m(t)\,u_m(t+\tau)}$$

If, in addition, only signal amplification is required, i. e.,

$$h(t+\tau)=au_c(t+\tau),$$

then

$$R_{yh}(\tau)=aR_c(\tau). \tag{2.13}$$

Solution of equation (2.10) leads to the following expression for the optimum linear filter transfer function $K(j\omega)$ [12]:

$$K(j\omega)=\frac{1}{2\pi\psi(j\omega)}\int_0^\infty e^{-j\omega t}\,dt\int_{-\infty}^{\infty}\frac{S_{yh}(\omega)}{\psi(-j\omega)}e^{j\omega t}\,d\omega, \tag{2.14}$$

where /33

$$|\psi(j\omega)|^2=S_y(\omega). \tag{2.15}$$

Here, $S_{yh}(\omega)$ and $S_y(\omega)$ -- power-density spectra corresponding to correlation functions $R_{yh}(\tau)$ and $R_y(\tau)$, i. e.,

$$\left.\begin{aligned} S_{yh}(\omega) &= \int\limits_{-\infty}^{\infty} R_{yh}(\tau)\, e^{-j\omega\tau}\, d\tau, \\ S_y(\omega) &= \int\limits_{-\infty}^{\infty} R_y(\tau)\, e^{-j\omega\tau}\, d\tau. \end{aligned}\right\} \qquad (2.16)$$

Since correlation functions $R_{yh}(\tau)$ and $R_y(\tau)$ are assumed as known during solution of the problem, then power spectra $S_{yh}(\omega)$ and $S_y(\omega)$ also are known. Therefore, formula (2.14) computations require only determination of complex magnitude $\Psi(j\omega)$, and modulus $|\psi(j\omega)|$, which preliminarily will be found from relationship (2.15).

The identical link to that between line system transfer function $K(j\omega)$ and its modulus $|K(j\omega)|$ exists between functions $\psi(j\omega)$ and $\psi(-j\omega)$. Therefore, the same methodology may be used for computations of $\psi(j\omega)$ with respect to the magnitude $|\Psi(j\omega)|$ found. This methodology is presented in § 18.3.

The minimum mean square error determined by the following formula corresponds to the transfer function determined by relationship (2.14)

$$\overline{\varepsilon^2_{min}} = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} [S_h(\omega) - |K(j\omega)|^2 S_y(\omega)]\, d\omega, \qquad (2.17)$$

where

$$S_h(\omega) = \int\limits_{-\infty}^{\infty} R_h(\tau)\, e^{-j\omega\tau}\, d\tau$$

is the function h(t) power-density spectrum.

Practical computations with formula (2.14) turn out to be rather enormous, while the formula itself is slightly unclear. A more convenient formula replacing (2.14) is presented in § 18.3. However, the computations here also remain enormous (see § 18.3). They are simplified significantly if the filter is not tasked with physical realization, (2.4), i. e., to assume that the lower limit in formula (2.3a) and expressions stemming from it equals $-\infty$ rather than zero.

Here, as Bode and Shannon demonstrated for the first time, optimum filter transfer function $K(j\omega)$ is determined by the following simple relationship [12]:

$$K(j\omega) = \frac{S_{yh}(\omega)}{S_y(\omega)}, \tag{2.18}$$

In a case where signal and noise are statistically independent, but /34
h(t) = u_c(t) (simple reproduction), this formula takes the form

$$K(j\omega) = \frac{S_c(\omega)}{S_c(\omega) + S_m(\omega)}, \tag{2.19}$$

where $S_c(\omega)$ and $S_m(\omega)$ -- power-density spectra of signal u_c(t) and noise $u_m(t)$, respectively.

Here, the mean square error equals

$$\overline{\varepsilon^2_{\text{мнк}}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_c(\omega)\, S_m(\omega)\, d\omega}{S_c(\omega) + S_m(\omega)}, \tag{2.20}$$

where $\varepsilon(t) = u_c(t) - \gamma(t)$ -- absolute error.

Relative error is characterized by ratio

$$\frac{\overline{\varepsilon^2_{\text{мнк}}}}{\overline{u_c^2}} = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{S_c(\omega)\, S_m(\omega)\, d\omega}{S_c(\omega) + S_m(\omega)}}{\displaystyle\int_{-\infty}^{\infty} S_c(\omega)\, d\omega}, \tag{2.20a}$$

where $\overline{u_c^2}$ -- signal voltage mean square. Although relationships (2.19) and (2.20) correspond to an optimum filter not realized physically, they are very useful since any physically-realized filter may not provide a mean square error less than (2.20).

Actually, assigning the filter a condition of physical realization narrows

the possibility of selecting the optimum filter characteristic and, as a result, may only worsen rather than improve the final result.

The optimum linear filter determined by relationship (2.10) or (2.14) is referred to as a Wiener filter and corresponds to a case where signal $u_c(t)$, noise $u_m(t)$, and synthesized filter are stationary. In future, the task of optimum linear filtration was generalized for the case of a transitory signal, noise, and filter. Here, as was the case with Wiener, it was assumed that signal and noise have zero mean values and known correlation functions.

In this more general case, optimum linear filter impulse response $\eta(t, \tau)$ is determined by integral equation

$$\int_{t_0}^{t} R_y(\tau, s)\, \eta(t, s)\, ds = R_{yh}(t, \tau), \quad t_0 \leqslant \tau \leqslant t. \tag{2.21}$$

If impulse response $\eta(t, \tau)$ satisfies this equation, then the mean (with respect to realizations) square error, $\overline{\varepsilon^2(t)}$, turns out at each moment in time to be minimal and equals

$$\overline{\varepsilon^2_{\text{мин}}} = \overline{h^2(t)} - \int_{t_0}^{t} \eta(t, \tau)\, R_{yh}(t, \tau)\, d\tau. \tag{2.22}$$

In formula (2.21), functions $R_y(\tau, s)$ and $R_{yh}(t, \tau)$, respectively, are /35
correlation functions $R_y(t_1, t_2)$ of input combination y(t) and cross-correlation functions $R_{yh}(t_1, t_2)$ between y(t) and required output effect h(t) from which the corresponding precise definition of the arguments is derived (for example, $R_{yh}(t, \tau)$ is obtained from $R_{yh}(t_1, t_2)$, with t replacing t_1 and τ replacing t_2).

Since processes y(t) and h(t) in this case may be transitory, then, as opposed to equation (2.10), the correlation functions will depend on t and τ, and not only on τ. In addition, since the synthesized system may turn out to be transitory, its impulse response equals $\eta(t, \tau)$, rather than $\eta(t)$ or $\eta(t - \tau)$. The ranges of integration in equation (2.21) are changed compared to (2.10) since, in this case, it is proposed that observation begin at moment t_0 and, consequently, considers a transitory process caused by connection at moment t_0 of the system itself or onset of input mixture y(t) at that moment, which is the same thing.

In the Wiener case examined previously, all processes in the system are stationary, i. e., it is assumed that $t_0 = -\infty$ for finite value t (or it is assumed that t_0 strives towards infinity when $t_0 = 0$, which is the same thing). It is not difficult to be convinced that equation (2.10) may be obtained from (2.21) as a partial case occurring when $t_0 = -\infty$,

$$R_{yh}(t_1, t_2) = R_{yh}(t_1 - t_2),$$
$$R_y(t_1, t_2) = R_y(t_1 - t_2) \quad \text{и} \quad \eta(t, \tau) = \eta(t - \tau).$$

Consequently, equation (2.21) may be considered as a generalized (for a stationary case) Wiener formula.

If noise $u_m(t)$ has the form of white noise and there is simple message reproduction, then

$$R_m(t_1, t_2) = S_{m0}\,\delta(t_1 - t_2)$$

and

$$R_{yh}(t_1, t_2) = R_c(t_1, t_2),$$

where S_{m0} -- noise spectral density (bilateral).

Here, as it is easy to see, equation (2.21) acquires the following form:

$$\int_{t_0}^{t} R_c(\tau, s)\,\eta(t, s)\,ds + S_{m0}\,\eta(t, \tau) = R_c(t, \tau), \quad t_0 \leqslant \tau \leqslant t. \qquad (2.21a)$$

In the general case, there is no success in finding the precise analytical solution for integral equation (2.21). At the present time, it is obtained only for a case where $u_c(t)$ and $u_m(t)$ are stationary random processes (see [127], for instance). There is success in obtaining a precise solution of equation (2.21) for transitory $u_c(t)$ and $u_m(t)$ only for several special types of correlated functions $R_y(t, \tau)$ and $R_{yh}(t, \tau)$. Thus, the method Shinbrot presented in [194] makes

it possible to obtain a precise solution if the correlation functions can be presented in the following form:

$$R_y(t,\tau)=\sum_{q=1}^{Q}a_q(t)\,b_q(\tau) \quad \text{with} \quad \tau\leqslant t,$$

$$R_y(t,\tau)=\sum_{q=1}^{Q}a_q(\tau)\,b_q(t) \quad \text{with} \quad \tau>t$$

and

$$R_{yh}(t,\tau)=\sum_{q=1}^{Q}c_q(t)\,b_q(\tau),$$

in which regard $W=\sum_{q=1}^{Q}[a_q(t)\,b_q(\tau)-a_q(\tau)\,b_q(t)]=W(t-\tau)$, i.e., it depends only on the difference $(t-\tau)$.

It should be noted that the theory of optimum linear filtration continues to develop continously. In recent years a large contribution to the development of this theory was made by r. Kolman. In the Kolman formulation the characteristics of the signal $u_c(t)$ [or the totality of signals $\overrightarrow{u_c(t)}$] are not given directly. Only the differential equation which it satisfies is known.

Such a more general and complex formulation of the filtration problem proves to be especially useful in the case of accomplishing optimum linear filtration in complex radio control spectrum. Here the structure of the optimum filter also proves to be complex and often can be practically realized only on the basis of digital electronic computers. A brief presentation of R. Kolman's theory is given, for example, in the book by R. Lee [153].

2.3. Linear Filters Which Ensure Maximum Signal/Noise Ratio.

Let the sum of signal and noise:

$$y(t)=u_c(t)+u_m(t),$$

operate on the input of a linear filter (Fig. 2.1) where is white noise described by relationshipis (1.19a, b, c, d), and is a signal with known amplitude spectrum , so that

$$u_c(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_c(j\omega)\,e^{j\omega t}\,d\omega. \tag{2.23}$$

Since the filter is a linear system, oscillation $\gamma(t)$ at filter output equals:

$$\gamma(t)=u_{c\,\text{вых}}(t)+u_{ш\,\text{вых}}(t), \tag{2.24}$$

where components $u_{c\,\text{вых}}(t)$ and $u_{ш\,\text{вых}}(t)$ are caused by the action of signal $u_c(t)$ and noise $u_{ш}(t)$, respectively.

If $K(j\omega)$ -- filter transfer function, then /37

$$u_{c\,\text{вых}}(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} K(j\omega)\,S(j\omega)\,e^{j\omega t}\,d\omega, \tag{2.25}$$

$$U_{ш}^2=\overline{u_{ш\,\text{вых}}^2(t)}=\frac{N_0}{2\pi}\int_0^{\infty}|K(j\omega)|^2\,d\omega. \tag{2.26}$$

Let t_0 be some fixed moment in time in which voltage $u_{c\,\text{вых}}(t)$ attains value $u_{c\,\text{вых}}(t_0)$. Then, in accordance with (2.25)

$$u_{c\,\text{вых}}(t_0)=\frac{1}{2\pi}\int_{-\infty}^{\infty} K(j\omega)\,S(j\omega)\,e^{j\omega t_0}\,d\omega. \tag{2.27}$$

We will designate

$$r=\frac{u_{c\,\text{вых}}(t_0)}{U_{ш}}. \tag{2.28}$$

It is evident that r is the ratio of signal output voltage (at moment t_0) to noise output voltage mean square value or, briefly stated, the ratio of signal (at moment t_0) to noise at filter output. The filter in which this ratio is the greatest is considered optimum.

From (2.25)--(2.28) it follows that

$$r=\frac{\dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{\infty} K(j\omega)\,S(j\omega)\,e^{j\omega t_0}\,d\omega}{\sqrt{\dfrac{N_0}{2\pi}\displaystyle\int_0^{\infty}|K(j\omega)|^2\,d\omega}}, \tag{2.29}$$

therefore, mathematically the problem boils down to finding that form of filter transfer function $K(j\omega)$ in which expression (2.29) attains the highest value. North solved this problem in 1943 using calculus of variables approaches [113]. In 1946, Van Vleck and D. Middleton independently solved the same problem with the aid of the Bunyakovskiy-Shvarts inequality (this method is presented in [8], for instance).

The result of this solution is that $r = r_{\text{макс}}$, if

$$K(j\omega) = aS^*(j\omega)\, e^{-j\omega t_0}, \tag{2.30}$$

where a -- arbitrary constant coefficient, while

$$S^*(j\omega) = S(-j\omega) \tag{2.31}$$

is a magnitude, complex conjugate to signal spectrum $S(j\omega)$ or, briefly stated, a complex conjugate signal spectrum.

Here /38

$$r^2 = r^2_{\text{макс}} = \frac{1}{\pi N_0}\int\limits_{-\infty}^{\infty} |S(j\omega)|^2\, d\omega. \tag{2.32}$$

But

$$\frac{1}{2\pi}\int\limits_{-\infty}^{\infty} |S(j\omega)|^2\, d\omega = \int\limits_{-\infty}^{\infty} u_c^2(t)\, dt = Q, \tag{2.33}$$

where Q -- signal energy at filter input.

Therefore, (2.32) also may be written in the following form:

$$r_{\text{макс}} = \left[\frac{u_{\text{с вых}}(t_0)}{U_{\text{ш}}}\right]_{\text{макс}} = \sqrt{\frac{2Q}{N_0}}. \tag{2.34}$$

It follows from this expression that magnitude $r_{\text{макс}}$, equalling ratio

$\left|\frac{u_{c\,вых}(t_0)}{U_{ш}}\right|_{макс}$, will not depend on t_0 if filter transfer function $K(j\omega)$ is selected for each new t_0 value in accordance with (2.30).

In other words, we may insure maximum signal-to-noise ratio determined by formula (2.34) for any predetermined moment in time t_0. All this requires is that filter transfer function $K(j\omega)$ be selected from formula (2.30), in which t_0 is included in the form of factor $e^{-j\omega t_0}$, accomplishing an output voltage shift over time by magnitude t_0. This result is fully understandable since, due to the noise being stationary, a shift over time in noise voltage does not change its mean square $U_{ш}^2$, while a shift over time of signal voltage displaces without distortions signal output voltage over time by the required magnitude.

It also follows from this that voltage $u_{c\,вых}(t)$ attains its maximum (peak) value at moment t_0 since, otherwise, it would be possible during a certain additional shift over time to obtain a greater signal-to-noise ratio value than follows from (2.34), which is impossible.

Thus, the optimum filter provides the maximum peak signal voltage value ratio to the noise voltage mean square value.

It follows from (2.30) that

$$|K(j\omega)| = a\,|S(j\omega)|, \tag{2.35}$$

i. e., the optimum filter frequency characteristic coincides (precise to a constant factor) with the signal amplitude spectrum. Therefore, this optimum filter often is referred to as a matched (with the signal) filter.

In order to solve the problem of the possibility of physical realization of a (2.30)-type filter transfer function, we will find impulse transient characteristic $\eta(t)$ corresponding to it.

Substituting (2.30) and (2.31) in (2.2), we obtain /39

$$\eta(t) = \frac{a}{2\pi}\int_{-\infty}^{\infty} S(-j\omega)\, e^{j\omega(t-t_0)}\, d\omega = \frac{a}{2\pi}\int_{-\infty}^{\infty} S(j\omega)\, e^{j\omega(t_0-t)}\, d\omega. \tag{2.36}$$

On the other hand, it follows from (2.23) that

$$u_c(t_0-t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} S(j\omega)\, e^{j\omega(t_0-t)}\, d\omega.$$

Comparing this expression with (2.36), we obtain the following expression for the optimum filter impulse transient characteristic:

$$\eta(t)=au_c'(t_0-t). \tag{2.37}$$

Since the physical realization condition has the form

$$\eta(t)=0 \quad \text{where} \quad t<0,$$

then meeting this condition requires that

$$u_c(t_0-t)=0 \quad \text{where} \quad t<0,$$

i. e.,

$$u_c(t)=0 \quad \text{where} \quad t>t_0 \tag{2.38}$$

Consequently, the optimum filter is physically realizable only when condition (2.38) is met. This condition signifies that signal $u_c(t)$ voltage at filter input must disappear prior to moment t_0 when signal voltage at filter output attains its maximum. It is possible to meet this condition for many, but not for all, signal $u_c(t)$ types.

Actually, let the signal at filter input damp with respect to exponent E_0e^{-bt}, for example. This signal disappears only when $t \to \infty$; thus, realization of condition (2.38) requires that voltage at filter output attain its maximum later than at infinity, which evidently is impossible. Consequently, t optimum filter is physically unrealizable for the signal depicted in Figure 2.2a.

However, in many actual cases, the signal at filter input damps sufficiently

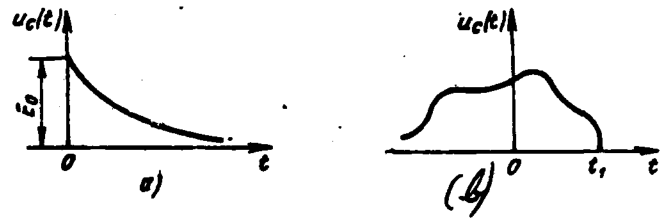


Figure 2.2

rapidly and it is possible to realize the optimal filter. Actually, let the /40 signal at filter input disappear at some moment t_1 (Figure 2.2b). Then, the filter will be physically realizable if we select

$$t_0 \geqslant t_1 \tag{2.39}$$

Consequently, when determining optimum filter transfer constant $K(j\omega)$ from formula (2.30), one may select any magnitude t_0, generally speaking, given that it satisfies relationship (2.39).

The selection usually is

$$t_0 = t_1 \tag{2.40}$$

for the following reason. Voltage $u_{c\,\text{вых}}(t)$ created by the signal at filter output attains maximum (peak) value at moment t_0 (Figure 2.3b). It usually is desirable for this maximum value to be achieved as early as possible, i. e., without additional delays over time. The lowest possible t_0 magnitude should be selected for this purpose, i. e., equal to moment t_1 of signal cessation at filter input.

On those rare occasions when, for some reason or another, it is desirable to obtain maximum signal value at output with slight time delay Δt rather than at moment t_1, value t_0 must be selected from condition

$$t_0 = t_1 + \Delta t. \tag{2.41}$$

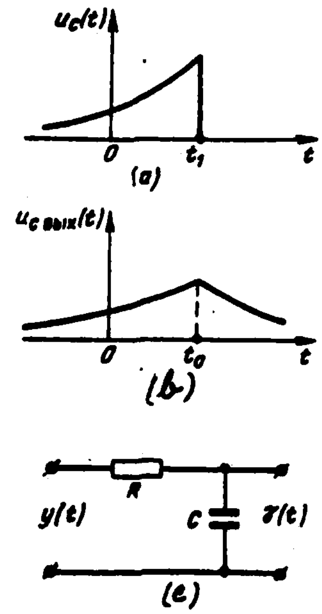


Figure 2.3

We note the following in conclusion. It was assumed during derivation of formula (2.34) that extant noise value U_m at filter output is determined in the steady-state mode, i. e., when noise appears at input at moment t_n, where $t_n = -\infty$. However, it was demonstrated in [174] that noise dispersion at matched filter output attains its steady-state value as early as time T following appearance of noise at its output (here T — signal $u_c(t)$ duration with which the given filter is matched). Therefore, if a matched filter is connected at the moment of signal appearance at output rather than at moment $t_n = -\infty$, formula (2.34) remains valid.

We will examine several examples.

Example 1 [8]. A signal has the form depicted in Figure 2.3a, to wit:

$$u_c(t) = \begin{cases} \frac{1}{RC} e^{(t-t_0)/RC} & \text{where } t < t_1, \\ 0 & \text{where } t > t_1. \end{cases}$$

Such a signal has a complex frequency spectrum /41

$$S(j\omega) = \frac{1}{1 - j\omega CR} e^{-j\omega t_1}.$$

Substitution of this formula into (2.30) leads (where a = 1 and $t_o = t_1$) to the following expression for optimal filter transfer function:

$$K(j\omega) = \frac{1}{1 + j\omega CR}.$$

Consequently, in this case the optimal filter will comprise cell R-C depicted in Figure 2.3c. Voltage $u_{e\,max}(t)$ at this filter's output has the form depicted in Figure 2.3b and is described by equations

$$u_{e\,max}(t) = \begin{cases} \frac{1}{2CR} e^{-(t_0 - t)/CR} & \text{where } t < t_0; \\ \frac{-1}{2CR} e^{-(t - t_0)/CR} & \text{where } t > t_0. \end{cases}$$

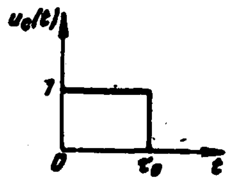


Figure 2.4

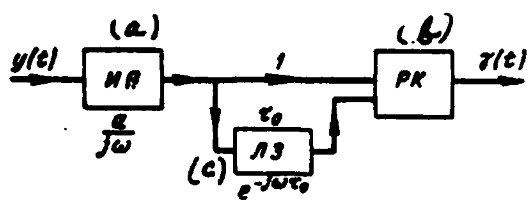


Figure 2.5. (a) -- IYa [integrating cell]; (b) -- RK [differential stage]; (c) -- LZ [delay line].

Example 2. A signal comprises a single square pulse with duration t_o disappearing at moment $t_1 = \tau_o$ (Figure 2.4).

Corresponding to this signal is complex spectrum

$$S(j\omega) = \frac{1}{j\omega}(1 - e^{-j\omega\tau_0}) \tag{2.42}$$

and, in accordance with (2.30) and (2.31), the optimum filter must have transfer function

$$K(j\omega) = \frac{a}{j\omega}\left(e^{j\omega\tau_0} - 1\right)e^{-j\omega t_0}. \tag{2.43}$$

In accordance with (2.40), we take $t_0 = t_1 = \tau_0$; then (2.43) takes the form

$$K(j\omega) = \frac{a}{j\omega}\left(1 - e^{-j\omega\tau_0}\right). \tag{2.44}$$

Such a transfer function is realized in the system depicted in Figure 2.5. Integrating cell IYa, having transfer constant $a/j\omega$, forms the first factor in formula (2.44). Delay line LZ, creating a delay by time τ_0 in combination with differential stages RK, has transfer constant $(1 - e^{-j\omega\tau_0})$, corresponding to the second factor in expression (2.44).

Optimum filter impulse transient characteristic $\eta(t)$ is determined by formula (2.37) and has the form depicted in Figure 2.6a for the signal being examined (Figure 2.4).

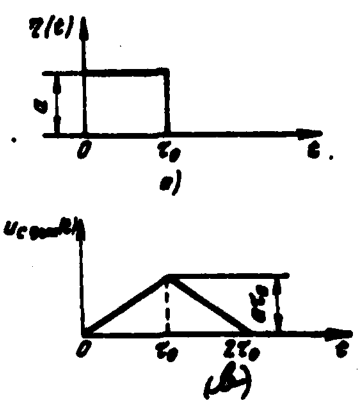


Figure 2.6

During practical determination of function $\eta(t)$ type, instead of (2.37), /42
it is often more convenient to use the following relationship stemming from it

$$\eta\left(\frac{t_0}{2}+\Delta t\right)=au_c\left(\frac{t_0}{2}-\Delta t\right), \tag{2.45}$$

i. e., impulse transient characteristic $\eta(t)$ is (precise to constant factor a) a mirror image of function $u_c(t)$ relative to point $t_0/2$.

In the specific example examined (Figure 2.4), when $t_0 = \tau_0$ and function $u_c(t)$ is symmetrical relative to point $t_0/2$, the mirror image coincides with the original and function $\eta(t)$ has (precise to a constant factor) a form identical to that of signal $u_c(t)$ voltage (Figure 2.6a).

Filter output voltage $u_{c\,вых}(t)$ equals

$$u_{c\,вых}(t)=\int_{-\infty}^{\infty}u_c(t-\tau)\,\eta(\tau)\,d\tau. \tag{2.46}$$

Considering that functions $u_c(t)$ and $\eta(t)$ have the form depicted in Figure 2.4 and 2.6a, respectively, from (2.46) we will find that $u_{c\,вых}(t)$ has the form depicted in Figure 2.6b, i. e., the optimum filter converts a signal rectangular pulse into a triangular pulse with double duration.

Example 3. The signal is a radio pulse with a rectangular envelope (Figure 2.7). The pulse has duration τ_0 and frequency f_0 of high-frequency occupation. It is possible to solve this problem, with both a precise approach by direct

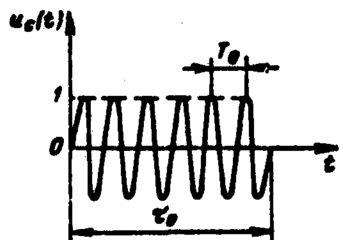


Figure 2.7

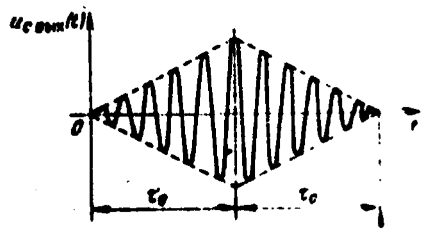


Figure 2.8

computation of function $K(j\omega)$ using formula (2.30) and approximately by the method of slowly-changing amplitudes.

One may demonstrate [71] through slowly-changing amplitudes that, if $K(j\omega)$

is the optimum transfer function for a video pulse, then $K[j(\omega-\omega_0)]$ will be the optimum transfer function for a radio pulse with the same envelope. Here, a filter output voltage envelope optimum for a radio pulse coincides with that optimum for a video pulse.

The resultant error here is less, the greater the ratio τ_0/T_0, the greater the high-frequency periods T_0 the radio pulses comprises. These assumptions are valid for a random-envelope radio pulse. The only requirement is for this envelope to exist, i. e., that the amplitude of the sinusoidal oscillation change slowly enough from one high-frequency period T_0 to another.

We will use the slowly-changing amplitudes approach for the Figure 2.7 pulse.

In light of all this, the optimal transfer function for this pulse may be obtained from expression (2.44) for a rectangular video pulse by replacing ω with $(\omega-\omega_0)$:

$$K(j\omega)=\frac{a}{j(\omega-\omega_0)}\left[1-e^{-j(\omega-\omega_0)\tau_0}\right]=$$
$$=\frac{a}{j(\omega-\omega_0)}\left(1-e^{j\omega_0\tau_0}e^{-j\omega\tau_0}\right). \tag{2.47}$$

For simplicity, we will assume that the radio pulse comprises a whole /43 number of periods, i. e., τ_0/T_0 = m, where m = whole number.

Then, (2.47) takes the form

$$K(j\omega)=\frac{a}{j(\omega-\omega_0)}\left(1-e^{-j\omega\tau_0}\right). \tag{2.47a}$$

It is possible to obtain a transfer function approximating $1/j(\omega-\omega_0)$ by using a highly-selective cavity resonator, while term $e^{-j\omega\tau_0}$ is obtained using a delay line with time delay τ_0.

However, as will be shown in 2.4, there is no need for a single radio pulse

when building such an optimum filter, since a much simpler quasi-optimum filter provides very similar results.

In light of this, an envelope at filter output optimum for a radio pulse coincides with filter output voltage optimum for a video pulse if the radio pulse at input has the same envelope as the video pulse. Thus, given the Figure 2.7 pulse, the pulse envelope at filter output has the same form as in Figure 2.6b, while the pulse itself has the form shown in Figure 2.8.

Consequently, an optimum filter converts a radio pulse with a rectangular envelope into one with a triangular envelope and double duration.

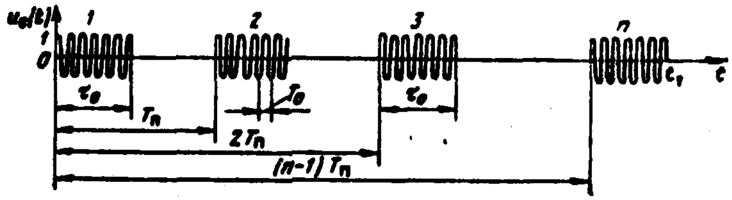


Figure 2.9

Example 4. A signal comprises train n of coherent radio pulses with spacing T_n , a rectangular envelope, and duration τ_0 (Figure 2.9).

We assume that

$$\frac{\tau_0}{T_0}=m; \quad \frac{T_n}{T_0}=l,$$

where T_0 -- high-frequency occupation; m and l -- whole numbers.

Computations provide this expression for the optimum filter complex /44 transfer constant:

$$K(j\omega)=K_1(j\omega)K_2(j\omega). \tag{2.48}$$

where $K_1(j\omega)$ -- filter complex transfer constant determined from formula (2.47a) optimum for a single radio pulse;

$$K_2(j\omega)=1+e^{-j\omega T_n}+e^{-j\omega 2T_n}+\ldots+e^{-j\omega(n-1)T_n}. \tag{2.49}$$

As usual, we assumed that $t_0 = t_1$ when deriving formulas (2.48) and (2.49).

It follows from Figure 2.9 that, in this case

$$t_0=t_1'=(n-1)T_n+\tau_0.$$

The Figure 2.10 optimum filter schematic diagram corresponds to formulas

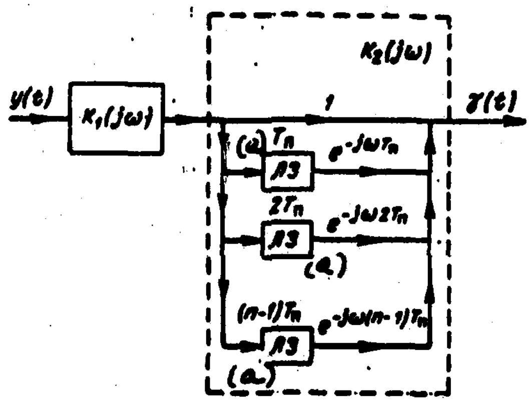


Figure 2.10. (a) -- LZ [delay line].

(2.48) and (2.49) and depicts filter $K_1(j\omega)$, optimum for a single radio pulse (see Example 3) and a delay line set.

It is evident that this delay line set may be replaced by one delay line

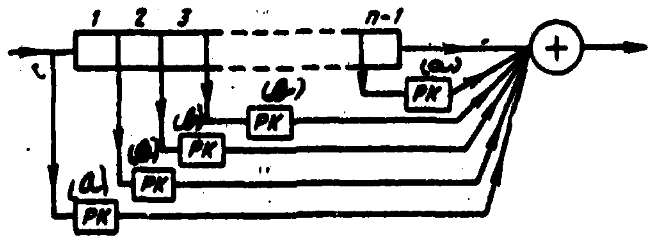


Figure 2.11. (a) -- RK [isolating stage].

with n--1 taps and isolating stages RK (Figure 2.11).

A large number of pulses n requires too many lines and taps. However, when $n \gg 1$, expression (2.49) may be simplified as follows.

We will designate

$$q = e^{-j\omega T_n}.$$

Then, $K_э(j\omega) = 1 + q + q^2 + \dots + q^{n-1}$. Where $n \rightarrow \infty$, we obtain /45

$$K_э(j\omega) = \frac{1}{1-q} = \frac{1}{1-e^{-j\omega T_n}}. \tag{2.50}$$

This transfer constant may be obtained in the Figure 2.12 circuit and comprises

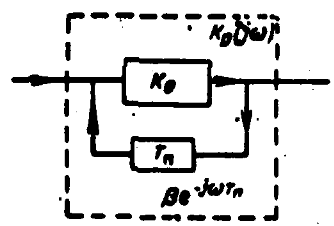


Figure 2.12

broadband amplifier with amplification K_0 enveloped by negative feedback across a delay line.

This circuit's transfer constant equals

$$K_p(j\omega) = \frac{K_0}{1 - K_0 \beta e^{-j\omega T_r}}. \tag{2.51}$$

Where

$$K_0 \beta = 1 \tag{2.52}$$

transfer constant $K_p(j\omega)$ coincides with required function $K_2(j\omega)$ (in principle, precision up to a constant factor does not play a role).

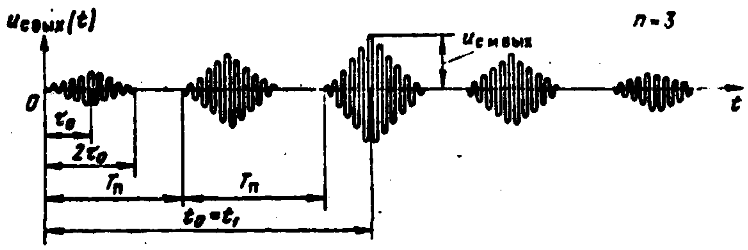


Figure 2.13

Optimum filter output voltage $u_{c\,вых}(t)$ has the form depicted in Figure 2.13 (where it is assumed for simplicity that n = 3).

It follows from (2.34) that signal-to-noise ratio $r^2_{макс}$ at filter output is proportional to signal energy Q:

$$r^2_{макс}=\frac{2Q}{N_0}. \tag{2.53}$$

For the signal shown in Fig. 2.9 $Q=nQ_1$, where Q_1 -- energy of one pulse, while Q -- energy of the entire signal (pulse "packet") comprising n periodic pulses.

Consequently,

$$r^2_{макс}=n\frac{2Q_1}{N_0}. \tag{2.54}$$

We now will explain what frequency characteristic the Figure 2.10 optimum filter and its links $K_1(j\omega)$ and $K_2(j\omega)$ have.

It follows from (2.35) that optimum filter frequency characteristic $|K(j\omega)|$ coincides with respect to form with signal frequency spectrum $|S(j\omega)|$. Therefore, frequency characteristic $|K(j\omega)|$ for a single radio pulse with a rectangular envelope has the form depicted in Figure 2.14a, while that for /46 an infinite periodic train of such pulses has the form depicted in Figure 2.14b.

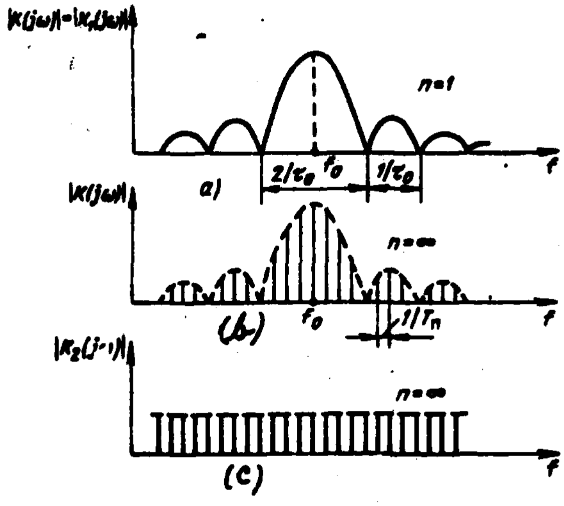


Figure 2.14

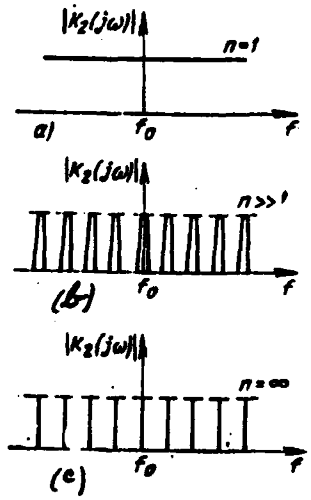


Figure 2.15

Here, the envelope of the Figure 2.14b characteristic (dotted curve) coincides with the characteristic for a single pulse.

It follows from (2.48) that

$$|K(j\omega)| = |K_1(j\omega)| \cdot |K_2(j\omega)|, \tag{2.55}$$

where $|K_1(j\omega)|$ — optimum filter frequency characteristic where n = 1, i. e., the Figure 2.14a characteristic.

It follows from formula (2.55) and Figure 2.14a, b that frequency characteristic $|K_2(j\omega)|$ where $n \to \infty$ must have the form depicted in Figure 2.14c. Actually, if you remultiply the ordinates of the curves in Figure 2.14a and b, then characteristic $|K(j\omega)|$, depicted in Figure 2.14b will be obtained. Consequently, the frequency characteristic of optimum filter unit $K_2(j\omega)$ (Figure 2.10) had, where $n \to \infty$, the form of a comb with infinitely-narrow teeth. One also may come to that conclusion through direct analysis of expression (2.49).

If the number of packet pulses n is great, but finite, then signal frequency spectrum $S(j\omega)$ ceases being linear, while frequency characteristic $|K_2(j\omega)|$ already will comprise teeth of finite length (Figure 2.15b) rather than infinitely-narrow teeth (Figure 2.14c). The smaller the number of pulses n, the greater the width of the comb teeth and, where n = 1, characteristic $|K_2(j\omega)|$ is converted into a horizontal straight line (Figure 2.15a).

From this example, it is possible to draw the following conclusions about the structure of a filter optimum for a "packet" (train) of n periodic pulses (Figure 2.9):

1. The filter will comprise two units -- $K_1(j\omega)$ and $K_2(j\omega)$.

2. Unit $K_1(j\omega)$ is a filter optimum for a single pulse. Therefore, this unit's frequency characteristic $|K_1(j\omega)|$ has the form of a continuous curve (Figure 2.14a)and will not depend on the number of pulses n.

3. Unit $K_2(j\omega)$ structure and frequency characteristic will depend considerably on the number of pulses n. Where $n \gg 1$, this unit has a comb-shaped frequency characteristic with narrow teeth (Figure 2.15b) and therefore is referred /47 to as a comb filter. The greater the n, the narrower the comb teeth.

4. A comb filter may be realized using various methods (see 2.10--2.12).

In the event a circuit with feedback is used (Figure 2.12), feedback magnitude βK_0 must be selected from the following relationships:

a) where $n = \infty$, it must be $\beta K_0 = 1$ (see formula (2.51)). However, this case is unrealistic since $n = \infty$ cannot occur and it is impossible to obtain stable operation of the Figure 2.12 circuit where $\beta K_0 = 1$;

b) Where n = 1, βK_0 must equal 0 (since $K_2(j\omega) = 1$ must be the case here);

c) The greater the n, the narrower the teeth must be of curve $|K_2(j\omega)|$ and, accordingly, the closer magnitude βK_0 is to unity.

However, it should be kept in mind that, where $n \neq \infty$, the Figure 2.12 filter with feedback does not correspond fully to an optimum comb filter (since expression (2.50) precisely coincides with (2.49) only where $n \to \infty$).

Example 5. The signal will comprise a rectangular video pulse train with spacing T_n (Figure 2.16). In this event, it is not difficult to obtain the following

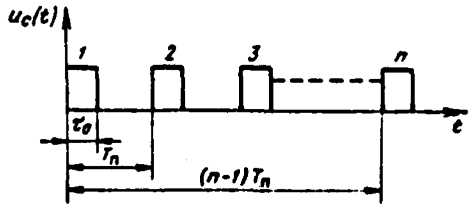


Figure 2.16

results, by means of both direct computation from formula (2.30) and the method of slowly-changing amplitudes.

An optimum filter, as was the case in the preceding example, will comprise two units $K_1(j\omega)$ and $K_2(j\omega)$ (Figure 2.10), where $K_1(j\omega)$ -- filter optimum for a single video pulse (see formula (2.44) and Figure 2.5), while $K_2(j\omega)$ -- comb filter described by formula (2.49).

Consequently, unit $K_2(j\omega)$ has a structure identical to that for the video pulse train examined in Example 4 and may be realized, in principle, by the identical means (Figures 2-10--2.12). However, comb filter realization in the case of radio pulses is significantly more complex than is the case for video pulses since, in the former case, much higher requirements are levied for precise accomplishment and time delay T_n stability.

If you denote tolerable time delay instability (or imprecise accomplishment) by ΔT_n, then for radio pulses, there is the requirement to meet the condition

$$\Delta T_n \ll T_0 = \frac{1}{f_0},$$

then, as was the case for video pulses, it is sufficient to meet the condition

$$\Delta T_n \ll \tau_0.$$

Actually, it follows from formula (2.49) that an optimum filter unit must

sum the pulses arriving at its input with time delays $T_n, 2T_n, \ldots, (n-1)T_n$. In order that the peak signal voltage obtained during summing is maximum, for radio pulses it is necessary that the time delay error be slight compared with pulse high-frequency occupation period T_0. This error must be as great as possible in the case of video pulses.

It also follows from formula (2.49) that, in essence, pulse accumulation be the basic operation during optimum periodic pulse train filtration. Such accumulation may be accomplished not only by means of delay lines (or line), but also by other known methods, such as with the aid of storage tubes.

2.4 Quasi-Optimum Linear Filters

Those linear filters, the shape of whose frequency characteristic is preassigned and maximum signal-to-noise ratio is provided only by appropriate frequency characteristic bandwidth selection, are referred to as quasi-optimum filters. In other words, a quasi-optimum filter is one in which only the bandwidth, rather than the frequency characteristic shape, is optimum.

The problem of finding such a filter was solved for the first time by V. I. Siforov [35]. Siforov examined a signal in the form of a single radio pulse with

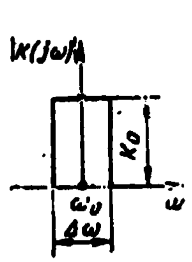


Figure 2.17

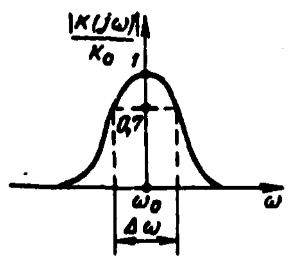


Figure 2.18

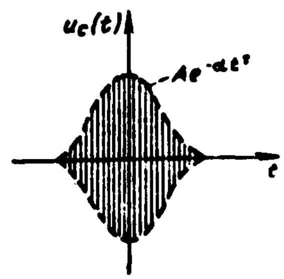


Figure 2.19

a rectangular envelope (Figure 2.7) and assumed that frequency characteristic $|K(j\omega)|$ has a rectangular shape with bandwidth $\Delta\omega = 2\pi\Pi$ (Figure 2.17). He demonstrated

that maximum signal-to-noise ratio r at the output of such a filter (formula 2.28) is obtained where the filter bandwidth value is

$$\Pi_{\text{опт}} = \frac{1{,}37}{\tau_0}. \tag{2.56}$$

Here

$$r^2_{\text{макс}} = 0{,}82\frac{2Q}{N_0}. \tag{2.57}$$

It follows from comparison of relationships (2.34) and (2.57) that, in the case of a quasi-optimum filter, signal energy Q greater by a factor of 1.22 is required to obtain such a value $r_{\text{макс}}$. Consequently, in this case a quasi-optimum filter provides a relatively small loss in required signal energy--less than 1 dB. Also important is the fact that the filter bandwidth P value turns out to be relatively slightly critical--when bandwidth P deviates by not more than a factor of 1.5 from its optimal value $\Pi_{\text{опт}}$ (towards an increase or decrease), magnitude r^2 decreases by not more than a factor of 1.25, i. e., the loss in required signal energy does not exceed 1 dB.

The frequency characteristic of actual filters has a smooth, rather than a rectangular, nature (Figure 2.18) and often, with respect to shape, approximates a gaussian curve of the form

$$\frac{|K(j\omega)|}{K_0} \approx e^{-a(\omega-\omega_0)^2}. \tag{2.58}$$

Here, for a radio pulse with a rectangular envelope, a quasi-optimum filter provides even better results than in the case of a rectangular frequency characteristic.

Actually, in the case of a radio pulse with a rectangular envelope, optimum frequency characteristic shape takes the form depicted in Figure 2.14a. It is evident from comparing Figures 2.14a, 2.17, and 2.18 that, in shape, an actual

frequency characteristic is closer to optimum than is a rectangular frequency characteristic.

An actual radio pulse has a smoother, rather than a strictly rectangular, envelope, which, in several instances, is shaped approximately like a gaussian curve of the type (Figure 2.19)

$$U_m(t) = Ae^{-\alpha t^2}. \tag{2.59}$$

This circumstance brings a quasi-optimum filter even closer to an optimum filter.

Actually, a pulse of the type (2.59) has frequency spectrum

$$|S(j\omega)| = Be^{-\frac{(\omega-\omega_0)^2}{4\alpha}}. \tag{2.60}$$

Therefore, frequency characteristic (2.58) is optimum for such a filter (given appropriate factor a selection, i. e., the width of the gaussian curve). Consequently, if the shape of a quasi-optimum filter frequency characteristic approximates a gaussian curve, while the radio pulse envelope also approximates a gaussian curve, then a quasi-optimum filter essentially coincides with an optimum filter.

It follows from this that, for actual radio pulses and actual frequency characteristics, a quasi-optimum filter provides even less loss compared to an optimaum filter than is the case examined above for the rectangular envelope and rectangular frequency characteristic, i. e., the resultant loss in required signal energy is significantly less than 1 dB. In addition, the deviation of bandwidth P from its optimal value in the case of real pulses and real frequency characteristics manifests itself considerably less than in the case examined above.

Thus, in the case of a single radio pulse, a quasi-optimum filter provides results almost identical to those for an optimum filter and there is no requirement to build strictly optimum filters. Selection of quasi-optimum filter bandwith is not very critical here. Evidently, these conclusions also are valid for a single video pulse.

In the case of a periodic radio pulse (or video pulse) train, the loss /50 when a quasi-optimum rather than an optimum filter is used is great and the magnitude of this loss increases with an increase in the number n of train pulses.

The magnitude of the loss also will strive towards infinity when $n \to \infty$.

Actually, when $n \to \infty$, the optimal frequency characteristic will comprise discrete infinitely-fine lines (Figure 2.14b) at a time when the frequency characteristic of a quasi-optimum filter (Figure 2.18) is continuous. Therefore, the effective area of the frequency characteristic square $|K(j\omega)|^2$ determining the power of the noise at output is an infinite number of times greater for a quasi-optimal frequency characteristic than for an optimum filter frequency characteristic.

The smaller the number n of train pulses, the greater the area of the optimal frequency characteristic teeth (Figure 2.15) and, consequently, the less the difference between optimal and quasi-optimal frequency characteristics.

In the case of a periodic pulse train, an optimum filter will, in essence, comprise two units--filter $K_1(j\omega)$, optimum for a single pulse and comb filter $K_2(j\omega)$ (Figure 2.10). It follows from what has been said that replacement of unit $K_1(j\omega)$ by a quasi-optimum filter in a majority of actual cases is permissible since it will not lead to significant signal-to-noise ratio loss.

Replacement of comb filter $K_2(j\omega)$ by a quasi-optimum filter (with a continuous frequency characteristic) leads to significant loss, radically rising with an increase in the number n of train pulses.

2.5 Comments on Communications and Differences Between Optimum Filters and Optimum Receivers

In conclusion, some comments are required concerning communications and differences in formulation of the problem of the optimum filter and the optimum receiver.

The problem posed will concern best reproduction of message x in the case of the optimum receiver.

As a modulated signal passes through a receiver, oscillation $u_c(t)$ in the ideal case and in the absence of noise also must be converted into message x, i. e., materially change its shape. Consequently, in the case of a receiver, even a very powerful change ("distortion") in signal $u_x(t)$ shape still does not give witness to distortions of message x this signal carries. Thus, for example, if the signal comprises a train of short-duration pulses modulated with respect to amplitude by low-frequency message x(t), then such a signal has a very broad spectrum and, during passage through a relatively-narrowband resonant amplifier /51
of high frequency, signal $u_x(t)$ shape will be distorted very strongly. However, the law of pulse amplitude modulation will remain essentially unchanged here and, following detection, may reproduce message x(t) almost without distortions. It then follows that, in the case of modulated signals (i. e., in the case where there is a requirement for best possible reproduction of only one or several signal parameters, rather than the entire signal), the optimum linear filters examined in §2.2 of this chapter may not best solve the problem, and sometimes even are completely unacceptable.

We will compare the following two cases as an illustration.

1. The requirement is to reproduce signal $u_x(t)$ with minimum mean square error.

2. The requirement is to reproduce message x carried by signal $u_x(t)$ with minimum mean square error.

In both instances, let the signal comprise a train of amplitude-modulated short-duration radio pulses, while noise is additive normal white noise.

We assume that noise intensity is slight compared with signal intensity. Then, in the first case, when any change in signal $u_x(t)$ shape is a distortion, it is advisible to use a filter with a very broad bandwidth.

In the second case, as will be shown in § 9.1 and elsewhere, the optimum

filter often boils down to a linear filter and subsequent detector. Here, the optimum linear filter is one providing maximum signal-to-noise ratio at detector input, i. e., a filter of the type examined in § 2.3. Such a filter has bandwidth $\Pi \approx 1/\tau_0$, i. e., significantly narrower than in the preceding case, and it causes very strong signal $u_x(t)$ shape distortions (for example, it converts radio pulses with a rectangular envelope into those with a triangular envelope, as shown in § 2.3). However, in spite of this, mean square error in message x(t) reproduction may be very slight.

Consequently, in the case examined, the filter providing the greater mean square error in signal $u_x(t)$ reproduction makes it possible to obtain a lesser mean square error in reproduction of the message this signal carries. For this reason, for the theory of optimal methods of reception of modulated signals, linear filters providing maximum signal-to-noise ratio at their output have a significantly greater value than linear filters providing minimum mean square error in signal reproduction at output.

Linear filters providing minimum mean square error are useful when the requirement is for best reproduction on a noise background of all signal $u_c(t)$ voltage, rather than one or several signal parameters. Therefore, they mainly may be used for radio receiving devices in stages connected beyond the detector.

Consequently, in radio receiving devices, filters providing maximum /52 signal-to-noise ratio mainly are used in the high-frequency amplifier (i. e., ahead of the demodulator), while filters providing minimum mean square error are used in the low-frequency amplfier (i. e., beyond the demodulator). The advisibility of including such filters in the optimum receiver is substantiated in more detail in subsequent chapters.

CHAPTER THREE /52

METHOD OF REDUCING "NON-WHITE" NOISE TO "WHITE" NOISE*

3.1 General Relationships

The method of "white" noise reduction, i. e., noise with random power spectrum $S_m(\omega)$, to "non-white" noise turns out to be useful in the theory of optimal reception methods. This method calls for signal-plus-noise y(t) preliminarily to be converted so that non-white noise within it is converted to white noise. The advisibility of doing so is tied in with the fact that finding the optimal system for white noise usually is a significantly easier problem. In addition, its solution (for white noise) in many cases already is known and the new problem thus boils down to one already solved.

V. A. Kotel'nikov proposed and used a method for reducing "non-white" noise to "white" noise for the first time in 1946 to solve the optimum receiver problem [1]. Subsequently, it also was used in several other works [5, 8, and others]. We will now examine this method.

*"Non-white" noise often is referred to also as correlated noise.

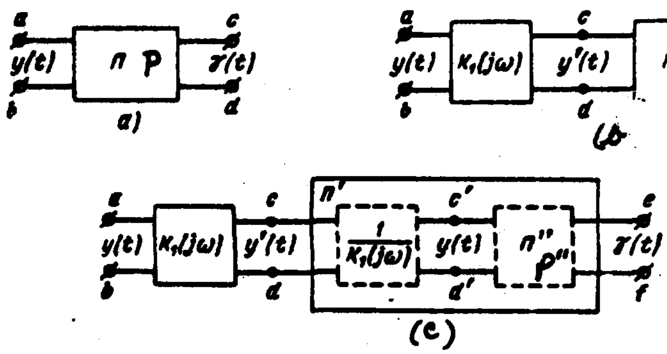


Figure 3.1

Let a combination of signal $u_x(t)$ and noise $u_m(t)$ arrive at receiver P input (Figure 3.1a):

$$y(t) = u_x(t) + u_m(t), \tag{3.1}$$

where $u_m(t)$ -- non-white noise having known power spectrum $S_m(\omega)$. The requirement is to construct receiver P in such a way that signal $u_x(t)$ (or message x) will be extracted in the best (in a certain sense) manner.

The following artificial routine may be used instead of direct solution of the problem. We will assume that combination y(t) passes across a linear four-terminal network with transfer constant $K_1(j\omega)$ (Figure 3.1b) and converts combination y(t) into y'(t) so that the non-white noise is converted to white noise:

$$y'(t) = u'_x(t) + u'_m(t), \tag{3.2}$$

where $u_m'(t)$ -- white noise, i. e., its power spectrum has the form

$$S_m'(\omega) = \text{const}.$$

For this, transfer constant $K_1(j\omega)$ must meet the condition

$$|K_1(j\omega)|^2 = \frac{b}{S_m(\omega)}, \tag{3.3}$$

where b -- random constant. Thus, combination $y'(t)$ of converted signal $u'_x(t)$ and white noise reach receiver P' input. Receiver P' selected is optimum, i. e., is such that it will extract signal $u'_x(t)$ (or message x) in the best way in the identical sense that receiver P did in the Figure 3.1a circuit.

We will demonstrate that the Figure 3.1b system, comprising linear filter $K_1(j\omega)$ and optimum receiver P', also insures identical quality extraction of signal $u_x(t)$ (or message x) from combination y(t), as did Figure 3.1a optimum receiver P.

The Figure 3.1b circuit differs from that in Figure 3.1a by presence at input of a linear four-terminal network with transfer constant $K_1(j\omega)$. It seems at first glance that inclusion of this network equates to placing a significant constraint on the optimum system found and must, as a result, lead to a worse result. However, in actuality, no such constraint will exist if you assume that desired optimum receiver P' may include a unit with transfer constant $1/K_1(j\omega)$ (Figure 3.1c).

Actually, in this event, initial oscillation y(t) is reestablished completely at receiver P" input and receiver P" may provide results identical to those of receiver P in the initial Figure 3.1a circuit.

Thus, we have shown that the Figure 3.1b system in principle may provide results identical to those the Figure 3.1a circuit provides.

We now will demonstrate that the Figure 3.1b system in principle not /54 may, but actually does, provide such results. While finding the optimum receiver P structure (Figure 3.1a), we start from the optimization predetermined mathematically from the minimum mean-square-error criterion, for example:

$$\overline{\delta^2} = \overline{[\gamma(t) - x(t)]^2} = \min,$$

where x(t) -- reproduced message real value, while $\gamma(t)$ -- result obtained at receiver output. While finding optimum receiver P' (Figure 3.1b), we will start with the exact same optimization (evidently, inclusion of compensating four-terminal network with transfer constant $K_1(j\omega)$ does not affect the form of this criterion).

We will assume that, as a result of mathematical analysis of the structure of optimum receiver P', we learned that the Figure 3.1b system provides poorer results (greater mean-square-error value $\overline{\delta^2}$, in this case) than did the Figure 3.1a system. This denotes that receiver P' provides poorer results (greater $\overline{\delta^2}$ magnitude at its output) than it may provide <u>in principle</u> (even though it was shown above that it in principle <u>may</u> provide just as good results as does receiver P). But, this contradicts receiver P' determination of optimality, in accordance with which its structure must be such that, of all its possible structures, it will provide the best possible results at output, i. e., minimum $\overline{\delta^2}$ magnitude.

Therefore, if when finding the optimum receiver P' structure we start with the optimization used for the initial system (Figure 3.1a) and we do not preclude this possible structure from containing unit $1/K_1(j\omega)$, then the Figure 3.1b system not only may, but also must, provide the same result as does the Figure 3.1a system.*

Thus, we have demonstrated that the Figure 3.1b system provides the same result as does the Figure 3.1a system. This denotes that the system optimum for non-white noise with spectrum $S_m(\omega)$ comprises a linear four-terminal network with transfer constant $K_1(j\omega)$ satisfying relationship (3.3) and receiver P' optimum for combination y'(t) comprising converted signal $u_x'(t)$ and white noise. Therefore, if the methodology for finding the receiver optimum for signal-plus-noise is known, then the Figure 3.1b circuit makes it possible to find the receiver optimum for non-white noise.

Extraction of "corrective" unit $K_1(j\omega)$ (Figure 3.1a, b) from the receiver P complement is a purely <u>intermediate</u> mathematical operation introduced only to simplify mathematical analysis. Therefore, there is no requirement for the /55 physical realization of unit $K_1(j\omega)$; it suffices for only the receiver as a whole to be physically realizable (or approximating physical realization).

This denotes that, if receiver P' (Figure 3.1b) comprises a linear system

*As noted in [175, page 212], such a result, strictly speaking, is completely accurate only assuming unrestricted processing time since, otherwise, complete "whitening" of the noise is impossible.

with transfer constant $K_2(j\omega)$ in the input portion, then there is a requirement for physical realization only of function $K_1(j\omega)K_2(j\omega)$, but not of each of its factors $K_1(j\omega)$ and $K_2(j\omega)$.

3.2 Using General Relationships to Find an Optimum Linear Filter

Let combination y(t) comprise signal $u_c(t)$ with known complex spectrum $S(j\omega)$ and noise $u_m(t)$ with power spectrum $S_m(\omega)$. Then, unit $K_1(j\omega)$ (Figure 3.1b) has a frequency characteristic determined by formula (3.3) and combination y'(t) will comprise signal $u_c'(t)$ and white noise $u_m'(t)$.

Signal $u_c'(t)$ complex spectrum $S'(j\omega)$ equals

$$S'(j\omega) = K_1(j\omega)S(j\omega), \tag{3.4}$$

while white noise $u_m'(t)$ has spectral density N_0', where $N_0' = \overline{(u_m')^2}/\Delta f$.

Let the requirement be to find a linear filter providing receipt of maximum ratio $r_{макс}$ of signal voltage to mean square noise voltage. This signifies that Figure 3.1b receiver P' in this case must be a linear filter providing maximum signal-to-noise ratio $r_{макс}$ for a signal with frequency spectrum $S'(j\omega)$ and white noise.

The solution to this problem was presented in § 2.3 and boils down to the fact that this filter's transfer constant $K_2(j\omega)$ must, in accordance with formula (2.30), meet the condition

$$K_2(j\omega) = aS'^*(j\omega)e^{-j\omega t_0}, \tag{3.5}$$

where $S'^*(j\omega)$ -- complex conjugate function with $S'(j\omega)$.

It follows from (3.4) that

$$S'^*(j\omega) = K_1^*(j\omega)S^*(j\omega). \tag{3.6}$$

$K_2(j\omega)$ -- this is the transfer constant of the Figure 3.1b P' unit.

Therefore, the entire optimum system's transfer constant $K(j\omega)$ equals

$$K(j\omega)=K_1(j\omega)K_2(j\omega). \tag{3.7}$$

It follows from (3.5)--(3-7) that

$$K(j\omega)=aK_1(j\omega)K_1^*(j\omega)S^*(j\omega)e^{-j\omega t_0}.$$

But,

$$K_1(j\omega)K_1^*(j\omega)=|K_1(j\omega)|^2=\frac{b}{S_m(\omega)};$$

consequently, /56

$$K(j\omega)=a_1\frac{S^*(j\omega)}{S_m(\omega)}e^{-j\omega t_0}, \tag{3.8}$$

where a_1 -- some constant magnitude.

Formula (3.8) also determines complex transfer constant $K(j\omega)$ of a linear filter insuring receipt of maximum signal-to-noise ratio in the case of non-white noise with power spectrum $S_m(\omega)$.

It follows from comparison of formulas (2.30) and (3.8) that, as opposed to the white noise formula, noise power spectrum $S_m(\omega)$ is included in the denominator of the non-white noise formula.

It follows from (3.8) that the optimum filter frequency characteristic is determined by the relationship

$$|K(j\omega)|=a_1\frac{|S(j\omega)|}{S_m(\omega)}, \tag{3.9}$$

i. e., this filter's frequency characteristic is proportional to the ratio of the signal ampl... spectrum to the noise power spectrum.

We will examine as our example a case where signal and noise spectra take the form of gaussian curves;

$$\left.\begin{aligned} |S(j\omega)|^2 &= S_0^2 e^{-(\omega-\omega_0)^2/2\Delta\omega_c^2}; \\ S_{\text{ш}}(\omega) &= S_{\text{ш}0} e^{-(\omega-\omega_0)^2/2\Delta\omega_n^2}. \end{aligned}\right\} \tag{3.10}$$

It is known that such spectra, although sufficiently close to physical realization, are not physically realizable.

It follows from (3.9) and (3.10) that the optimum filter's frequency characteristic takes the form

$$|K(j\omega)| = a_1 \frac{S_0^2}{S_{\text{ш}0}} \exp\left\{-\frac{(\omega-\omega_0)^2}{2\Delta\omega_c^2}\left[1-2\left(\frac{\Delta\omega_c}{\Delta\omega_n}\right)^2\right]\right\}, \tag{3.11}$$

where $\Delta\omega_c/\Delta\omega_n$ -- ratio of the width of the signal and noise spectra.

It follows from this formula that, given

$$\frac{\Delta\omega_c}{\Delta\omega_n} < \frac{1}{\sqrt{2}} \tag{3.12}$$

the optimum filter's frequency characteristic takes the form of a gaussian curve and, consequently, is close to being physically realizable. If condition (3.12) is not met, frequency characteristic $|K(j\omega)|$ is very remote from physical realization.

Now, we will explain what the energy Q' of the converted signal is in this case.

$$Q' = \frac{1}{2\pi} \int_{-\infty}^{\infty} |S'(j\omega)|^2 \, d\omega.$$

Considering (3.3) and (3.4), we have /57

$$Q' = \frac{b}{2\pi} \int_{-\infty}^{\infty} \frac{|S(j\omega)|^2}{S_m(\omega)} d\omega. \tag{3.13}$$

From relationships (3.10) and (3.13), we obtain

$$Q' = \frac{b}{\sqrt{2\pi}} \cdot \frac{S_0^2}{S_{m0}} \cdot \frac{\Delta\omega_c}{\sqrt{1 - \left(\frac{\Delta\omega_c}{\Delta\omega_n}\right)^2}}. \tag{3.14}$$

Where $\Delta\omega_c/\Delta\omega_n \geqslant 1$, energy Q' will become infinitely-great or imaginary and the solution found becomes senseless; therefore, it must be

$$\frac{\Delta\omega_c}{\Delta\omega_n} < 1. \tag{3.15}$$

However, it follows from comparison of relationships (3.12) and (3.15) that condition (3.12) is more rigid than (3.15). This signifies that the validity of the solution found is constrained by the fact that the solution found is very far from physical realization, rather than by the fact that energy Q' will become infinitely-great or imaginary.

Thus, in the case of spectra of the (3.10) type, optimum solution (3.11) found makes sense only when condition (3.12) is met.

If the noise spectrum is narrow in comparison with the signal, the solution found turns out to be inapplicable. However, in this case, the optimum solution is evident—it suffices to connect a linear trap with trap band equalling the noise spectrum width (the latter here is understood to mean the frequency region outside of which noise spectrum energy is infinitesimally-small) in order to obtain maximum signal-to-noise ratio at receiver output. Then, noise voltage essentially will be absent at filter output and signal-to-noise ratio will be very great.

3.3 Conclusions

The problem of optimum signal $u_x(t)$ (or message x this signal carries) reception on a background of additive non-white noise with power spectrum $S_m(\omega)$ boils down to finding receiver P' optimum (in the same sense) for converted signal $u_x'(t)$ and white noise.

The converted signal is obtained from the initial signal with passage through a linear filter with transfer constant $K_1(j\omega)$, whose modulus is determined by relationship

$$|K_1(j\omega)|^2 = \frac{b}{S_m(\omega)}.$$

Desired optimum receiver P for signal $u_x(t)$ will comprise receiver P' and linear filter with transfer constant $K_1(j\omega)$ preceding it.

Methods obtained for white noise using the aforementioned approach in many cases may be generalized relatively simply in the case of noise with an essentially-random power spectrum. Therefore, subsequent material is presented mainly as it relates to white noise.

PART TWO /59

OPTIMAL RECEPTION OF PRECISELY-KNOWN SIGNALS (KOTEL'NIKOV THEORY OF POTENTIAL NOISE IMMUNITY)

CHAPTER FOUR

GENERAL RELATIONSHIPS

4.1 Problem Formulation

Let the signal-noise sum reach receiver input (Figure 1.1)

$$y(t)=u_x(t)+u_m(t), \tag{4.1}$$

where $u_x(t)$ -- precisely-known signal, i. e., a signal whose only unknown parameter is de~ired message x, while $u_m(t)$ -- noise.

Message x at the point of reception is considered a random variable or random time function with known a priori distribution P(x). Noise distribution law $W_m(u_m)$ also is assumed to be known. The receiver analyzes oscillation y(t) during predetermined (finite) time interval T (Figure 4.1) and, based on this analysis, must reproduce message x in the best possible way (in the sense indicated below). Since the signal is distorted by random noise $u_m(t)$, while analysis time T is

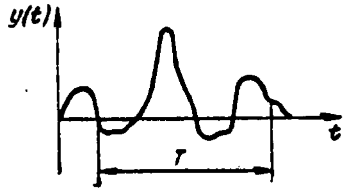


Figure 4.1

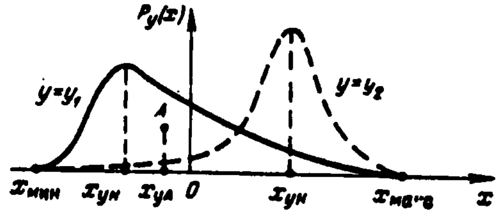


Figure 4.2

finite, then there is no receiver in principle that may reproduce message x completely accurately, with complete reliability. There always will be a certain probable error.

Consequently, the most one can expect of a receiver under such conditions is to determine the probability of a particular value (or of a particular realization) of message x for a given realization of total oscillation y(t) arriving at receiver input.

Mathematically, this denotes that, based on analysis of total oscillation y(t), the ideal receiver must compute distribution $P_y(x)$ for all possible values (or realizations) of message x for given realization y(t) (given realization y(t) always is understood to mean that form of oscillation y(t) which it has in given observation cycle T).

$P_y(x)$ is a distribution of a random variable if message x is a discrete /60 random magnitude; if x is an analog random magnitude (or a time function), then $P_y(x)$ is a probability density (see Figure 4.2, for example).

In future, for brevity we will call $P_y(x)$ a distribution of a random variable in all cases, understanding however that, in all cases where x is an analog random magnitude (or a time function), $P_y(x)$ is a probability density.

Distribution $P_y(x)$ is referred to as a posteriori (empirical) since it may be found only as a result of analysis of oscillation y(t) realization. Distribution

$P_y(x)$ differs in principle thereby from a priori distribution $P(x)$, which is assumed to be known beforehand, i. e., prior to analysis of oscillation $y(t)$.

Distribution $P_y(x)$ often is referred to also as an inverse probability distribution since it shows what the probabilities are of certain values of cause x if effect y brought about by this cause is known.

In future, for brevity we will call $P_y(x)$ an inverse probability distribution.

Thus, the best that a receiver in principle can offer on the basis of analysis of realization $y(t)$ is to compute distribution $P_y(x)$ of message x inverse probabilities for all possible values of these messages (see Figure 4.2, for example).

A decision then must be made from analysis of the distribution type as to what will be the value x of a transmitted message. This decision may be made by an operator (observing the Figure 4.2 $P_y(x)$ distribution on an oscillograph screen, for example) or can be made by the receiver itself. In the latter case, a rule must be made for the receiver so that it will make its decision γ based upon analysis of distribution $P_y(x)$. One such rule may be the so-called maximum inverse probability principle.

The assumption in this event is that

$$\gamma = x_{y\text{н}}, \tag{4.2}$$

where $x_{y\text{н}}$ -- most probable message value, i. e., that value at which inverse probability $P_y(x)$ has the greatest value (Figure 4.2). In the general /61 case, the shape of curve $P_y(x)$ for various realizations of $y(t)$ and, consequently, the most probable value, may differ (Figure 4.2). This circumstance is underscored by introduction of index y into designation $x_{y\text{н}}$.

Consequently, when the maximum inverse probability principle is used, the receiver issues that message value γ most probable for a given $y(t)$ realization.

Another decision rule also may be established for the receiver. For example,

it is possible to require the receiver to compute abcissa x_{yA} of the center of "gravity" of the area captured under curve $P_y(x)$ (Figure 4.2) and supply

$$\gamma = x_{yA}. \tag{4.3}$$

Kotel'nikov demonstrated that, given known conditions, such a rule insures minimum message reproduction mean square error.

It is possible, in the general case, to propose an even infinite number of rules advisible for this or that reason by which the receiver must make a decision. Each such rule may be considered an appropriate receiver optimization.

It will be shown in Chapter 19 that, in several instances, different rules will lead to identical results. Thus, for example, in certain instances, the maximum inverse probability principle also results in minimum mean square error, i. e., rules (4.2) and (4.3) may provide identical results.

However, in the general case, different optimum receiver structures and properties correspond to different optimizations.

Kotel'nikov considers optimum that receiver operating on the maximum inverse probability principle, i. e., in accordance with rule (4.2). This principle is one of the simplest and most natural and, in many cases, turns out to be most advisible.

4.2 Computation of Message Inverse Probabilities

Inverse probability $P_y(x)$ may be found from the following general relationships of the theory of probabilities:

$$P(x, y) = P(x) P_x(y) = P(y) P_y(x), \tag{4.4}$$

where $P(x, y)$ -- mutual probability of two random magnitudes (or functions) x and y; $P_x(y)$ -- conditional probability of y for a given x; $P(y)$ -- unconditional probability y.

It follows from (4.4) that

$$P_y(x) = \frac{1}{P(y)} P(x) P_x(y). \tag{4.5}$$

Since when computing function $P_y(x)$ we are interested in the relationship of this function to x when y is unchanged, factor 1/P(y) in expression (4.5) may be replaced by some constant k:

$$P_y(x) = kP(x) P_x(y). \tag{4.6}$$

Coefficient k may be determined from normality condition:

$$\int_{A_x} P_y(x)\,dx = 1, \tag{4.7}$$

where A_x — region of all possible values of x.

Therefore*

$$k = \frac{1}{\int_{A_x} P(x) P_x(y)\,dx}. \tag{4.8}$$

Since a priori distribution P(x) is known, then only distribution $P_x(y)$ remains to be found.

It follows from formula (4.6) that one must know the relationship of $P_x(y)$ to x for a given y in order to determine function $P_y(x)$. Distribution $P_x(y)$, considered a function of x for a given y, is referred to as a likelihood function.

$P_x(y)$ is the probability (probability density) of y for a given x. Since

---

*If x is a random time function, then expressions $\int_{A_x} P_y(x)\,dx$ and $\int_{A_x} P(x) P_x(y)\,dx$ are an abbreviated notation for integrals $\int_{A_x} \cdots \int P_y(x_1, \ldots, x_n)\,dx_1 \ldots dx_n$ and $\int_{A_x} \cdots \int P(x_1, \ldots, x_n) P_{x_1, \ldots, x_n}(y)\,dx_1 \ldots dx_n$, respectively.

function $u_x(t)$ is known precisely for a given x, then it follows from (4.1) that, for a given x, the probability of realization y(t) coincides with the probability of that realization of noise $u_{ш}(t)$ equalling difference $[y(t)-u_x(t)]$. But, the probability (probability density) of realization $u_{ш}(t)$ for given x where x and $u_{ш}$ are independent is characterized by noise $W_{ш}(u_{ш})$ probability density.

Therefore, if message x and noise $u_{ш}(t)$ statistically are independent (which usually is the case), then,

$$P_x(y)=W_{ш}(y-u_x). \tag{4.9}$$

Kotel'nikov demonstrated that noise has the form of normal white noise. The distribution of such noise, in accordance with (1.25), has the following form:

$$W_{ш}(u_{ш})=\frac{1}{(\sqrt{2\pi N})^n}e^{-\frac{1}{N_0}\int_0^T u_{ш}^2(t)\,dt}$$

Therefore /63

$$P_x(y)=\frac{1}{(\sqrt{2\pi N})^n}e^{-\frac{1}{N_0}\int_0^T [y(t)-u_x(t)]^2\,dt} \tag{4.10}$$

Substituting this expression into (4.6), we obtain

$$P_y(x)=k_1P(x)e^{-\frac{1}{N_0}\int_0^T [y(t)-u_x(t)]^2\,dt}, \tag{4.11}$$

where k_1 — constant coefficient, which will not depend on x and which may be found from normality condition (4.7).

A receiver optimum in this sense, as Kotel'nikov assumed, uses formula (4.11) to compute inverse probability $P_y(x)$ for all possible message x values and supplies in the form of decision γ that value of x at which function $P_y(x)$ has the greatest

value. It often is convenient to represent expression (4.11) in a slightly-different form for practical realization of such a computer. Formula (4.11) is squared in order to do this. Then we will obtain

$$P_y(x) = k_1 P(x) e^{-\frac{1}{N_0}\int_0^T y^2(t)\,dt} e^{-Q_x/N_0} e^{\xi_x}, \tag{4.12}$$

where

$$Q_x = \int_0^T u_x^2(t)\,dt \tag{4.13}$$

is the energy of the signal carrying message x, which in the general case may depend on x (for instance, for amplitude modulation); magnitude ξ_x is determined from relationship

$$\xi_x = \frac{2}{N_0}\int_0^T y(t)\,u_x(t)\,dt. \tag{4.14}$$

Factor $e^{-\frac{1}{N_0}\int_0^T y^2(t)\,dt}$ in expression (4.12) for a given y will not depend on x and therefore may be considered as constant magnitudes as probability $P_y(x)$ is computed. Consequently, we may write expression (4.12) in the form

$$P_y(x) = k_2 P(x) e^{-Q_x/N_0} e^{\xi_x}, \tag{4.15}$$

where k_2 -- constant factor, which may be computed from normality condition (4.7).

It follows from expression (4.15) that computation of ξ_x from formula /64 (4.14), i. e., finding the common correlation between received oscillation y(t) and anticipated signal $u_x(t)$, is the main operation when determining inverse probability $P_y(x)$.

4.3 Optimum Receiver Structure

The device computing the integral in expression (4.14) is called a correlator.

Consequently, the correlator is the main optimum receiver element. It may be realized, for example, using the Figure 4.3 schematic and comprises a multiplier

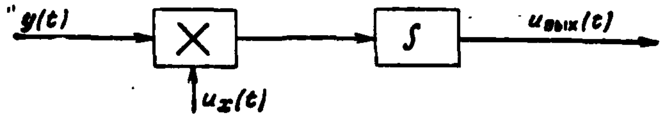


Figure 4.3

and an integrator. Here, output voltage equals

$$u_{\text{вых}}(t) = c\int_{t_1}^{t} y(t)\, u_x(t)\, dt, \tag{4.16}$$

where c -- proportionality factor, whose magnitude is determined by the properties of the actual physical devices used to carry out the multiplication and integration operations. It follows from (4.16) that this coefficient has dimension [1/(B times c)]. Expression (4.16) coincides with (4.14) precise to a constant coefficient and integration limits.

It evidently is not difficult to eliminate the difference in the constant coefficient by appropriate selection of the multiplier or integrator transfer constant.

There are several ways to obtain requisite integration limits, including connection of a model (copy) of anticipated signal $u_x(t)$ only for time interval 0 – T. Here, by moment t = T, output voltage will attain requisite value

$$u_{\text{вых}}(T) = c\int_{0}^{T} y(t)\, u_x(t)\, dt. \tag{4.17}$$

In future, this voltage will be retained unchanged (for the ideal integrator). Therefore, there is a requirement to bring the integrator output to zero prior to conducting a new test for message x extraction.

The linear filter described in § 2.3, matched with the anticipated signal, may be used instead of a correlator if signal $u_x(t)$ lasts a finite time /65
interval $(0 - T)$ (which usually is the case).

Actually, if oscillation $y(t)$ arrives at linear filter input, then the voltage at its output at moment t is

$$u_{вых}(t) = \int_{-\infty}^{t} y(z)\,\eta(t-z)\,dz,$$

where $\eta(t)$ — filter impulse transient characteristic.

If the filter is optimum for signal $u_x(t)$ (in the signal-to-noise sense), then its characteristic satisfies relationship (2.37):

$$\eta(t) = a u_x(t_0 - t).$$

Since, according to the assumption, signal $u_x(t)$ disappears at moment $t = T$, then one may assume $t_0 = T$, i. e.,

$$\eta(t) = a u_x(T - t).$$

Therefore, the oscillation at the output of such a filter equals;

$$u_{вых}(t) = a \int_{-\infty}^{t} y(z)\,u_x(T - t + z)\,dz.$$

Since, according to the assumption, there is no signal also when $t < 0$, then $u_x(T - t + z) = 0$ where $T - t + z < 0$, i. e., where $z < t - T$; therefore

$$u_{вых}(t) = a \int_{t-T}^{t} y(z)\,u_x(T - t + z)\,dz.$$

At moment $t = T$, we obtain

$$u_{вых}(T) = a \int_{0}^{T} y(z)\,u_x(z)\,dz. \qquad (4.18)$$

A comparison of expressions (4.17) and (4.18) shows that they coincide with precision to a constant factor. Consequently, an optimum linear filter matched with anticipated signal $u_x(t)$ may be used instead of a correlator (one should assume that $t_0 = T$ during synthesis of such a filter). Oscillation y(t) is supplied to filter input and the magnitude of oscillation $u_{вых}(t)$ at filter output at moment t = T is measured (clipped).

In spite of the fact that, given the conditions enumerated above, a matched filter and correlator at moment t = T supply identical results, they differ greatly where their remaining properties are concerned. In particular, when a correlator is used (Figure 4.3), there is the requirement at the point of reception to generate voltage $u_x(t)$ (signal copy) corresponding to the anticipated signal. As opposed to this, a matched filter is a passive system not requiring generation of /66 oscillation $u_x(t)$. It is simpler to realize a correlator for certain signal types (phase-shift, for instance), while a matched filter is simpler for others (phase-modulated, for example).

In several cases, it is advisible to use a combination of correlator and matched filter, i. e., to build a so-called correlation-filtration device. The possibility of such a device is based on the following [129].

It often is possible to represent signal $u_x(t)$ in the form of the product of two known time functions, i. e., to assume

$$u_x(t) = f_1(t) f_2(t). \tag{4.19}$$

For example, if the signal is a segment of a sinusoid with duration T, one may assume

$$\left.\begin{aligned} &f_1(t) = U\sin(\omega_0 t + \psi), \\ &f_2(t) = \begin{cases} 1, & \text{where } t = 0 - T, \\ 0, & \text{outside these limits} \end{cases} \end{aligned}\right\} \tag{4.20}$$

i. e., to consider that $f_1(t)$ -- sinusoid (of infinite duration), while $f_2(t)$ -- rectangular video pulse.

Substituting (4.19) into (4.14), we obtain

$$\xi_x = \frac{2}{N_0}\int_0^T y(t) f_1(t) f_2(t)\,dt. \tag{4.21}$$

This expression may be represented in the form

$$\xi_x = \frac{2}{N_0}\int_0^T y'(t) f_2(t)\,dt, \tag{4.22}$$

where

$$y'(t) = y(t) f_1(t). \tag{4.23}$$

But, based on the aforementioned, a transform of the (4.22) type may be accomplished (precise to a constant factor) by a linear filter matched with the "signal"

$$u_x'(t) = f_2(t), \tag{4.24}$$

while transform (4.23) may be accomplished by a multiplier. Therefore, operation (4.14) may be accomplished by the device depicted in Figure 4.4, where ОФ' --

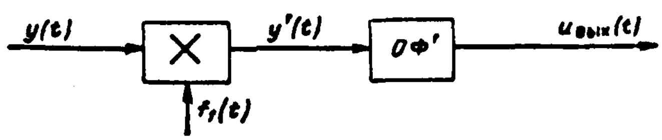


Figure 4.4

filter matched with signal $u_x'(t)$. At moment t = T, voltage at this device's output will be identical to that at the Figure 4.3 output and equalling (precise to a constant factor) the requisite ξ_x magnitude.

The Figure 4.4 device is referred to as a correlation-filtration device since it will comprise, along with a matched filter, a multiplier as a component part

of the correlator. Just like a correlator, the device requires generation at the point of reception of a "signal copy"--function $f_1(t)$. Since multipliers $f_1(t)$ and $f_2(t)$ usually are simpler functions than complete oscillation of /67 signal $u_x(t)$, use of a correlation-filtration device often makes it possible to simplify considerably the "signal copy" generator (compared to a purely correlation device) and matched filter (compared to a purely filtration device).

For instance, if signal $u_x(t)$ with duration T comprises a periodic train of unmodulated radio pulses and the selected envelope of this train is $f_1(t)$, then the Figure 4.4 schematic will have the following structure: "signal copy" $f_1(t)$ -- periodic rectangular video pulse train, while *ОФ'* -- linear filter matched with a sinusoidal signal with duration T.

The "signal copy" in the Figure 4.4 schematic will be a sinusoid, while filter *ОФ'* must be matched with the video pulse train if rf (sinusoidal) occupation, rather than radio pulse train envelope, is selected as function $f_1(t)$.

It follows from what has been said that computation of magnitude ξ_x for any one specific type of signal $u_x(t)$, i. e., for one message x value, does not present significant difficulties. However, to find distribution $P_y(x)$, magnitude ξ_x must be determined for all, rather than for one, possible message x values (or all possible realizations) which, in the general case, presents major difficulties. These difficulties are decreased considerably in the case of discrete messages, when x may have only a finite number m of discrete values. Therefore, in the case of analog messages, message digitization (quantization) often is used to simplify practical computational device realization.

It often is advisible to determine not function $P_y(x)$ itself, but its logarithm $\ln P_y(x)$ to simplify practical realization and theoretical computations.

It follows from (4.15) that

$$\ln P_y(x) = \ln k_2 + \ln P(x) - \frac{Q_x}{N_0} + \xi_x. \tag{4.25}$$

Since the greatest $\ln P_y(x)$ and $P_y(x)$ values coincide when x changes, the

optimum receiver operating on the maximum inverse probability principle must select that value of x for which $\ln P_y(x)$ has the greatest (in the algebraic sense) value.

Coefficient k_2 does not impact upon the form of curves (4.15) and (4.25); /68
therefore, for simplicity $\ln k_2$ may be discarded.

Then, instead of (4.25), we obtain

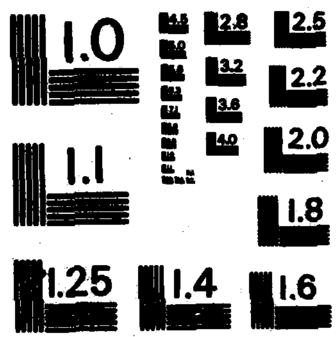
$$\ln P_y(x) = \ln P(x) - \frac{Q_x}{N_0} + \xi_x. \tag{4.25a}$$

Expression (4.25a) is significantly simpler than (4.15) for computations.

The resultant relationships are valid, both for discrete messages and for individual analog message values and for oscillations x(t). A more detailed examination of these three cases, in the order that Kotel'nikov approached them [1], will be provided in the following section.

AD-A120 899 THEORY OF OPTIMUM RADIO RECEPTION METHODS IN RANDOM NOISE(U) FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OH L S GUTKIN 24 SEP 82 FTD-ID(RS)T-0784-82 2/7

UNCLASSIFIED F/G 9/4 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

CHAPTER FIVE

RECEPTION OF DISCRETE MESSAGES

5.1 General Case

Let message x have one and only one of possible values

$$x_0,\ x_1,\ x_2,\ \ldots,\ x_k,\ \ldots,\ x_m,$$

where x_0 -- zero value corresponding to the absence of any message (spacing).

A priori probabilities of the presence (sample) of these messages equals, respectively

$$P(x_0),\ P(x_1),\ \ldots,\ P(x_k),\ \ldots,\ P(x_m),$$

while

$$\sum_{k=0}^{m} P(x_k) = 1. \qquad (5.1)$$

Each value x_k of the message has its corresponding value $u_{x_k}(t)$ of the signal.

Therefore, one and only one of the following possible signals must be present at receiver input during interval (0, T):

$$u_{x_0}(t),\ u_{x_1}(t),\ \dots,\ u_{x_k}(t),\ \dots,\ u_{x_m}(t). \tag{5.2}$$

In future, for simplicity magnitude x is dropped from the indices, i. e., instead of (5.2), we write

$$u_0(t),\ u_1(t),\ \dots,\ u_k(t),\ \dots u_m(t). \tag{5.2a}$$

Zero message ($x_0 \equiv 0$) may be transmitted, either by a zero signal (in this case, $u_0(t) \equiv 0$ — passive spacing), or by any predetermined known signal /69 (then $u_0 \not\equiv 0$ — active spacing); meanwhile it is known beforehand which of these two cases must occur.

Signals $u_k(t)$ are assumed to be precisely known, i. e., the shape of all signals of the (5.2a) type is known precisely. The only unknown subject to determination is exactly which of these signals was present at receiver input during interval (0, T).

Consequently, completely-determined signal $u_k(t)$ corresponds to each message x_k and, on the other hand, if signal $u_k(t)$ is known, then message x_k corresponding to it is determined unambiguously in the same way. Therefore, a priori probabilities $P(u_k)$ of the presence of signals $u_k(t)$ equal the a priori probabilities $P(x_k)$ of the corresponding messages:

$$P(u_k) = P(x_k). \tag{5.3}$$

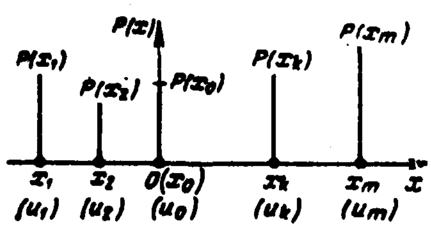


Figure 5.1

In this event, a priori distribution P(x) has the form depicted in Figure 5.1.

Using analysis of realization y(t) as its basis, the receiver must determine exactly which of the m + 1 possible values message x had in the interval (0, T) examined, i. e., which exactly of the m + 1 possible signals existed at that time at receiver input.

Since signal u_k determination thereby denotes determination also of message x_k corresponding to it, then for brevity we will discuss only signals in the future.

Here, general relationships (4.12)--(4.14), and (4.25a) may be written in the form

$$P_y(u_k) = k_3 P(u_k)\, e^{-Q_k/N_0} e^{\xi_k}; \qquad (5.4)$$

$$\ln P_y(u_k) = \ln P(u_k) - \frac{Q_k}{N_0} + \xi_k; \qquad (5.4a)$$

$$Q_k = \int_0^T u_k^2(t)\, dt; \qquad (5.5)$$

$$\xi_k = \frac{2}{N_0} \int_0^T y(t)\, u_k(t)\, dt. \qquad (5.6)$$

Since the receiver operates on the maximum inverse probability principle, then it reproduces each time that signal u_k for which inverse probability $P_y(u_k)$ turns out to be the greatest. Here, various kinds of errors are possible and composite error probability $P_{\text{ош}}$ equals

$$\begin{aligned} P_{\text{ош}} = {} & P(u_0) P_{u_0}(\neq u_0) + P(u_1) P_{u_1}(\neq u_1) + \dots + \\ & + P(u_k) P_{u_k}(\neq u_k) + \dots + P(u_m) P_{u_m}(\neq u_m). \end{aligned} \qquad (5.7)$$

where $P_{u_k}(\neq u_k)$ -- conditional probability that, given signal u_k at input, /70
the receiver will reproduce some other signal (it is unimportant which one), i. e., an error occurs.

It often happens to be simpler to compute composite correct reproduction probability $P_{\text{прав}}$, rather than composite error probability $P_{\text{ош}}$:

$$P_{\text{прав}}=P(u_0)\,P_{u_0}(u_0)+P(u_1)\,P_{u_1}(u_1)+\ldots+P(u_m)\,P_{u_m}(u_m), \tag{5.8}$$

where $P_{u_k}(u_k)$ -- probability (conditional) that the receiver will reproduce signal u_k when signal u_k actually is at input.

It is evident that

$$P_{\text{прав}}=1-P_{\text{ош}}. \tag{5.9}$$

It is easy to become convinced that a receiver operating on the maximum inverse probability principle provides a minimum composite error probability value.

Actually, $P_y(u_k)$ is the probability that signal u_k will occur for a given y.

Each time, the optimum receiver at output supplies that signal u_k, for which the probability is the greatest for a given y. Consequently, signal selection error probability for a given y is the least. This occurs for any y(t) realization.

Consequently, for <u>every</u> y(t) realization, the receiver supplies the response with the minimum error probability. Therefore, resultant composite error probability, considering all possible oscillation y(t) realizations, must also be the minimum possible here. Consequently, from the point of view of minimum composite error probability, the maximum inverse probability principle provides the best possible results. However, the minimum composite error probability criterion for discrete messages is not the only possible and advisible one.

Actually, given minimization of probability $P_{\text{ош}}$ (5.7), errors of all types uniformly are considered dangerous. For example, if signal u_2 was active at input, then reproduction of signal u_3 instead of u_2 is considered just as undesirable ("dangerous") as reproduction of signal u_7 or u_9 and so on. However, in several cases (in radar, for instance), different types of errors have a different "danger" and there is a need to insure the minimum, not of expression (5.7), but of a more-complex expression, which considers the "weight" ("danger") of certain errors.

In this case, appropriate corrections (examined in Part Four) must be introduced into the maximum inverse probability principle.

This chapter examines only the simplest case where a receiver operates on the maximum inverse probability principle in accordance with the work of Kotel'nikov.

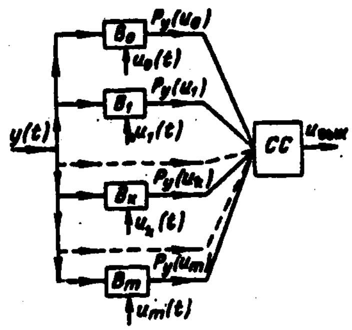


Figure 5.2

Here, the structure of the optimum receiver must, in accordance with formulas (5.4)--(5.6), take the form depicted in Figure 5.2.

The receiver will comprise computers $B_0, B_1, \ldots, B_m$, to the inputs of /71 which oscillation y(t) and anticipated signals $u_0, u_1, \ldots, u_m$ are supplied. Computed inverse probabilities $P_y(u_0), P_y(u_1), \ldots, P_y(u_m)$ are compared in comparison circuit CC and a determination is made as to which has the highest value; the signal (or message) corresponding to it also is supplied in the form of output magnitude $u_{\text{вых}}$.

Since inverse probabilities $P_y(u_k)$ always are positive, then they may be reflected by the corresponding values of positive constant voltage. Therefore, circuit CC may be realized, for instance, with the aid of a cathode-ray tube [CRT] so that a specific area on the CRT screen will correspond to each channel (each computer B_k). Here, the brightest area will be the one corresponding to the channel with the highest value $P_y(u_k)$. Then, selection of the brightest area suffices for signal u_k selection.

Each computer B_k may have the structure depicted in Figure 5.3, i. e., comprise a correlator computing ξ_k from formula (5.6), summing stage *CK* , and stage *ЭK*

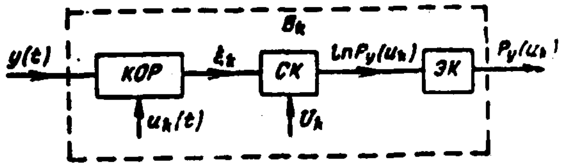


Figure 5.3

with an exponential characteristic, which converts ln $P_y(u_k)$ into $P_y(u_k)$; here, stage *ЭK* in this simplest case, when the task is only to determine which of the $P_y(u_k)$ values is the greatest, is not required in principle since this problem may be solved by comparison, not of the $P_y(u_k)$ magnitudes themselves, but through comparison of their logarithms, ln $P_y(u_k)$.

It is evident from Figure 5.3 that U_k magnitudes, depending on a priori probabilities $P(u_k)$ and anticipated signal energies Q_k, play the role of biases in the optimum receiver circuit.

5.2 Binary Detection

We now will examine in more detail the simplest case, called binary detection.

In this case, the signal may have only two values, one of which identical with zero, i. e.,

$$u_0(t) \equiv 0; \quad u_1(t) = u_c(t).$$

The problem boils down here, in essence, to signal detection, i. e., /72 to clarification of whether or not signal $u_c(t)$ (and noise) is present at receiver input or a signal is absent (i. e., only noise is present).

To solve the problem of whether or not there is a signal, the receiver must compare inverse probabilities $P_y(u_c)$ and $P_y(0)$, where, in accordance with formula (5.4),

$$\left.\begin{aligned} \ln P_y(u_c) &= \ln P(u_c) - \frac{Q}{N_0} + \xi, \\ \ln P_y(0) &= \ln P(0). \end{aligned}\right\} \tag{5.10}$$

Here $P(u_c)$ and $P(0)$ -- a priori probabilities of signal presence and absence, while Q -- signal energy.

The receiver must supply the answer "yes" (signal), if $\ln P_y(u_c) > \ln P_y(0)$, and the answer "no" (no signal), if $\ln P_y(u_c) \leqslant \ln P_y(0)$.

Consequently, the receiver must supply the answer "yes" when this inequality is satisfied

$$\xi > U_0, \tag{5.11a}$$

where

$$U_0 = \ln \frac{P(0)}{P(u_c)} + \frac{Q}{N_0}; \tag{5.11b}$$

$$\xi = \frac{2}{N_0} \int_0^T y(t)\, u_c(t)\, dt. \tag{5.11c}$$

If inequality (5.11a) is not satisfied, the receiver must supply the answer "no."

It follows from formulas (5.11a) and (5.11c) that the structure of the optimum receiver for binary detection has the form depicted in Figure 5.4. The correlator computes ξ from formula (5.11c); the ξ value found is compared at moment $t = T$ with threshold (constant negative bias) U_0. If the result is $\xi > U_0$,

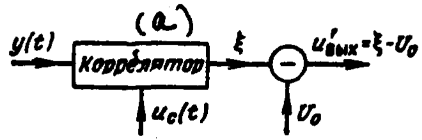


Figure 5.4. (a) -- Correlator.

i. e., $u_{вых}' > 0$, the receiver supplies the answer "yes" (signal). Otherwise, it supplies the answer "no" (no signal, only noise).

As noted in § 4.3, the correlator in several cases may be replaced by a matched linear filter or by a correlation-filtration device.*

In the Figure 5.4 circuit, correlator output and threshold bias are characterized by dimensionless magnitudes ξ and U_0 determined by relationships (5.11). However, in an actual circuit, correlator output voltage (or that of the matched linear /73
filter replacing it) at the moment of comparison with the threshold equals

$$u_{вых}(T) = c\int_0^T y(t)\, u_c(t)\, dt, \tag{5.12}$$

where c — some proportionality coefficient conveniently determined from the following circumstances. When noise is absent (or in the presence of a signal considerably more intense than the noise), correlator (or matched filter) output voltage at moment t = T equals

$$u_{c\,вых}(T) = c\int_0^T u_c(t)\, u_c(t)\, dt = cQ,$$

where Q — signal energy. Consequently,

*Technical advantages and shortcomings of replacing a correlator with a matched filter are examined in [136, 188, and others].

$$c = \frac{u_{c\,\text{вых}}(T)}{Q}. \qquad (5.13)$$

It follows from substitution of expressions (5.11) and (5.12) that threshold bias at actual circuit output must equal

$$U_{0p} = \frac{cN_0}{2} U_0 = u_{c\,\text{вых}}(T) \frac{N_0}{2Q} U_0. \qquad (5.14)$$

Sometimes, it is more convenient to determine parameters c and U_{op} from extant noise voltage $U_ш$ output value, rather then from usable signal output voltage $u_{c\,\text{вых}}(T)$.

Since in accordance with (2.34)

$$\frac{u_{c\,\text{вых}}}{U_ш} = \sqrt{\frac{2Q}{N_0}},$$

then, instead of (5.13) and (5.14), one may assume

$$c = \frac{\sqrt{2}\,U_ш}{\sqrt{N_0 Q}}; \qquad (5.13a)$$

$$U_{0p} = U_ш \sqrt{\frac{N_0}{2Q}}\, U_0. \qquad (5.14a)$$

It follows from (5.11b) and (5.14) that

$$U_{0p} = u_{c\,\text{вых}}(T)\left[\frac{1}{2} + \frac{N_0}{2Q} \ln \frac{P(0)}{P(u_c)}\right]. \qquad (5.15)$$

In particular, if P(0) = P(u_c), then

$$U_{0p} = \frac{1}{2} u_{c\max}(T),$$

i. e., threshold bias must equal one-half the maximum value of usable signal /74 output voltage.

Two types of errors are possible when this type detector is used: false alarms, i. e., "yes" when in actuality there is no signal, and signal misses, i. e., "no" when there actually is a signal.

We will designate the probabilities of false alarms and signal misses $P_{лт}$ and $P_{пр}$, respectively. Then, composite error probability equals

$$P_{ош} = P(0) P_{лт} + P(u_c) P_{пр}. \tag{5.16}$$

Since a priori probabilities P(0) and P(u_c) are assumed to be known, then it remains to compute conditional error probabilities $P_{лт}$ and $P_{пр}$.

Initially, we will find false alarm probability $P_{лт}$. A false alarm means "yes" when there is no signal. Therefore, the false alarm probability equals the probability of satisfying inequality (5.11a) when there is no signal, i. e., when

$$y(t) = u_ш(t).$$

It follows from (5.11c) that, in this case

$$\xi = \frac{2}{N_0} \int_0^T u_ш(t) u_c(t) dt, \tag{5.17}$$

and the false alarm probability is the probability that random magnitude ξ, determined from formula (5.17), will exceed threshold U_0.

We will find the law of magnitude ξ distribution. On the basis of formula (1.12), one may write

$$\left.\begin{aligned} u_{\text{m}}(t) &= \sum_{k=1}^{n} u_{\text{m}\,k}\,\psi_k(t), \\ u_{\text{c}}(t) &= \sum_{k=1}^{n} u_{\text{c}\,k}\,\psi_k(t). \end{aligned}\right\} \tag{5.18}$$

Substituting these expressions into (5.17) and considering relationship (1.12c), we obtain

$$\xi = \frac{2}{N_0}\,\frac{1}{2f_\text{B}} \sum_{k=1}^{n} u_{\text{m}\,k} u_{\text{c}\,k} = \sum_{k=1}^{n} u_{\text{m}\,k} \left(\frac{u_{\text{c}\,k}}{N}\right), \tag{5.19}$$

where N = $N_0 f_\text{B}$.

Since signal $u_c(t)$ is precisely known, then values u_{ck} included in formula (5.19) are not random magnitudes.

Ordinates $u_{\text{m}k}$ of white noise $u_{\text{m}}(t)$ statistically are independent random magnitudes with a normal law of distribution, zero average values, and dispersions equalling N (see § 1.3).

Therefore, terms $u_{\text{m}k}(u_{\text{c}k}/N)$ in the sum of (5.19) also are independent /75 random magnitudes with a normal law of distribution, zero average values, and dispersions equalling

$$N\left(\frac{u_{\text{c}\,k}}{N}\right)^2 = \frac{u_{\text{c}\,k}^2}{N}.$$

Consequently, magnitude ξ determined from formula (5.19) also has a normal law of distribution with zero average values and with dispersion σ_ξ^2 equalling the sum of the dispersions n of the terms, i. e.,

$$P(\xi) = \frac{1}{\sqrt{2\pi}\,\sigma_\xi} e^{-\xi^2/2\sigma_\xi^2}, \tag{5.20}$$

where

$$\sigma_\xi^2 = \frac{1}{N} \sum_{k=1}^{n} u_{ck}^2 .$$

But, it follows from (1.12d) that

$$\sum_{k=1}^{n} u_{ck}^2 = 2f_s \int_0^T u_c^2(t)\, dt = 2f_s Q;$$

therefore

$$\sigma_\xi^2 = \frac{2Q}{N_0} . \tag{5.21}$$

The probability that ξ will exceed threshold U_0 equals

$$P_{\pi\tau} = \frac{1}{\sqrt{2\pi}\sigma_\xi} \int_{U_0}^{\infty} e^{-\xi^2/2\sigma_\xi^2}\, d\xi,$$

i. e.,

$$P_{\pi\tau} = \frac{1}{\sqrt{2\pi}} \int_{\alpha_1}^{\infty} e^{-z^2/2}\, dz, \tag{5.22}$$

where

$$\alpha_1 = \frac{U_0}{\sigma_\xi} = \frac{\ln \dfrac{P(0)}{P(u_c)} + \dfrac{Q}{N_0}}{\sqrt{\dfrac{2Q}{N_0}}} . \tag{5.23}$$

We now will find signal miss probability P_{np} .

It follows from Figure 5.4 that there will be a signal miss in those cases when, given a signal at input, i. e., when

$$y(t) = u_c(t) + u_m(t), \tag{5.24}$$

it turns out that

$$\xi \leqslant U_0. \tag{5.25}$$

Therefore, miss probability is the probability that inequality (5.25) /76 is satisfied, in which ξ is determined by formulas (5.11c) and (5.24), while it is precisely that

$$\xi = \frac{2}{N_0}\int_0^T [u_c(t) + u_m(t)]\, u_c(t)\, dt.$$

Expanding the parentheses, we obtain

$$\xi = \frac{2}{N_0}\int_0^T u_m(t)\, u_c(t)\, dt + \frac{2Q}{N_0}. \tag{5.26}$$

The first term in this expression already has been investigated in this section [see formula (5.17)]; here we found that it has a normal law of distribution with a zero average value and dispersion $2Q/N_0$. Therefore, magnitude ξ, determined from relationship (5.26), has a normal law of distribution with average value $2Q/N_0$ and dispersion $2Q/N_0$ also, i. e., in this case

$$P(\xi) = \frac{1}{\sqrt{2\pi}\sigma_\xi} e^{-\left(\xi - \frac{2Q}{N_0}\right)^2 / 2\sigma_\xi^2}, \tag{5.27a}$$

where σ_ξ^2 is determined as usual from formula (5.21).

Therefore, the miss probability, i. e., the probability that $\xi \leqslant U_0$, equals

$$P_{np} = \frac{1}{\sqrt{2\pi}\sigma_\xi} \int_{-\infty}^{U_0} e^{-\left(\xi - \frac{2Q}{N}\right)^2 / 2\sigma_\xi^2} d\xi.$$

After appropriate replacement of variables, we obtain

$$P_{np} = \frac{1}{\sqrt{2\pi}} \int_{\alpha_0}^{\infty} e^{-z^2/2} dz, \tag{5.27b}$$

where

$$\alpha_3 = \frac{\frac{Q}{N_0} - \ln \frac{P(0)}{P(u_c)}}{\sqrt{\frac{2Q}{N_0}}}. \tag{5.27c}$$

It is possible to use formulas (5.16), (5.22), and (5.27b) to compute the dependence of error probabilities $P_{\text{ош}}$, $P_{\text{лт}}$ and P_{np} on ratio $P(0)/P(u_c)$ of the a priori probabilities and signal-to-noise ratio Q/N_0.

However, in those instances when permissible error probabilities are slight (do not exceed 0, 1), requisite signal-to-noise ratio Q/N_0 is so large that /77
it is possible to use an asymptotic expansion of the probability integral

$$V(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-z^2/2} dz, \tag{5.28a}$$

and precisely the expansion

$$\begin{aligned} V(x) &= \frac{1}{\sqrt{2\pi}} \frac{e^{-x^2/2}}{x} \left(1 - \frac{1}{x^2} + \frac{1 \cdot 3}{x^4} - \dots\right) \approx \\ &\approx \frac{1}{\sqrt{2\pi}} \cdot \frac{e^{-x^2/2}}{x}. \end{aligned} \tag{5.28b}$$

Here, formulas (5.22) and (5.27b) take on the following form:

$$P_{\text{лт}} \approx \frac{1}{\sqrt{2\pi}\,\alpha_1} e^{-\alpha_1^2/2}, \quad P_{\text{пр}} \approx \frac{1}{\sqrt{2\pi}\,\alpha_2} e^{-\alpha_2^2/2}; \tag{5.29}$$

hence

$$\left.\begin{aligned} \frac{\alpha_1^2}{2} &\approx \ln \frac{1}{P_{\text{лт}}} - 0{,}9 - \ln \alpha_1, \\ \frac{\alpha_2^2}{2} &\approx \ln \frac{1}{P_{\text{пр}}} - 0{,}9 - \ln \alpha_2. \end{aligned}\right\} \tag{5.30a}$$

If $P_{\text{лт}} \ll 1$ and $P_{\text{пр}} \ll 1$, it is possible to assume that

$$\left.\begin{aligned} \frac{\alpha_1^2}{2} &\approx \ln \frac{1}{P_{\text{лт}}}, \\ \frac{\alpha_2^2}{2} &\approx \ln \frac{1}{P_{\text{пр}}}, \end{aligned}\right\} \tag{5.30b}$$

and, considering expressions (5.23) and (5.27b), we obtain

$$q \approx \left(\sqrt{\ln \frac{1}{P_{\text{лт}}}} + \sqrt{\ln \frac{1}{P_{\text{пр}}}} \right)^2, \tag{5.31}$$

$$\frac{P_{\text{пр}}}{P_{\text{лт}}} \approx \frac{P(0)}{P(u_0)}. \tag{5.32}$$

Here and elsewhere, $q = Q/N_0$. Formulas (5.31) and (5.32) will become precise when

$$P_{\text{лт}} \to 0, \quad P_{\text{пр}} \to 0.$$

Due to the rise in magnitudes $P_{\text{лт}}$ and $P_{\text{пр}}$, the error of expressions (5.31)

and (5.32) increases. As computations show, when $P_{лт} \leq 0,1$ and $P_{пр} \leq 0,1$, instead of (5.31), it is more advisible to use the following corrected formula:

$$q \approx \left(\sqrt{\ln\frac{1}{P_{лт}} - 1,4} + \sqrt{\ln\frac{1}{P_{пр}} - 1,4}\right)^2. \qquad (5.31a)$$

The error in determination of q does not exceed 0.5 db if $P_{лт} \leqslant 0,1$ and $P_{пр} \leqslant 0,1$. Where $P_{лт} \to 0$ and $P_{пр} \to 0$, the error asymptotically will strive towards zero. It follows from formula (5.16) and (5.32) that /78

$$P_{лт} = \frac{P_{ош}}{2P(0)}, \quad P_{пр} = \frac{P_{ош}}{2P(u_c)}. \qquad (5.33)$$

If a priori probabilities P(0) and $P(u_c)$ and permissible composite error probability $P_{ош}$ are given, then it is possible to use formulas (5.33) to compute permissible probabilities $P_{лт}$ and $P_{пр}$ and then to use formula (5.31a) to determine requisite signal-to-noise ratio q.

According to problem conditions, a priori probabilities P(0) and $P(u_c)$ often are unknown. Here, formula (5.16) of composite probability $P_{ош}$ becomes senseless and should be provided, not by permissible magnitude $P_{ош}$, but by permissible conditional false alarm and signal miss probabilities $P_{лт}$ and $P_{пр}$. Here, requisite signal-to-noise ratio q also is determined from formula (5.31a).

During derivation of formula (5.31a), it was assumed that threshold bias U_o at optimum receiver output is determined from formula (5.11b), and it is precisely that

$$U_0 = \ln\frac{P(0)}{P(u_c)} + \frac{Q}{N_0}.$$

If a priori probabilities P(0) and $P(u_c)$ are unknown and we are given directly conditional probabilities $P_{лт}$ and $P_{пр}$, then, in accordance with expression (5.32), ratio $P(0)/P(u_c)$ in this formula should be replaced by $P_{пр}/P_{лт}$, i. e., one should assume

$$U_0 = \ln \frac{P_{np}}{P_{\pi T}} + q.$$

Here, formula (5.14a) takes on the following form:

$$U_{0p} = U_{m} \frac{q + \ln \frac{P_{np}}{P_{\pi T}}}{\sqrt{2q}}. \tag{5.34a}$$

However, for given probability $P_{\pi T}$, it is possible to obtain a simpler formula for determination of the requisite real threshold U_{0p}.

Actually, in absence of a signal, voltage at moment t = T at correlator output (or at the output of the matched filter replacing it) has a normal law of distribution with a zero expected value and dispersion equalling U_m^2. Therefore,

$$P_{\pi T} = \frac{1}{\sqrt{2\pi} U_m} \int_{U_{0p}}^{\infty} e^{-u_{\text{вых}}^2(T)/2U_m^2} du_{\text{вых}}(T) = \\ = \frac{1}{\sqrt{2\pi}} \int_{U_{0p}/U_m}^{\infty} e^{-z^2/2} dz.$$

Where $P_{\pi T} < 0{,}1$ (this usually is the case), it is possible with sufficient /79 precision to assume

$$\frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-z^2/2} dz \approx \frac{1}{\sqrt{2\pi}\alpha} e^{-\alpha^2/2},$$

where $\alpha = U_{0p}/U_m$. Hence, it follows that

$$U_{0p} \approx U_m \sqrt{2 \ln \frac{1}{P_{\pi T}}} \tag{5.34b}$$

(such a result, of course, is obtained also from formula (5.34a) if expression (5.31) is substituted into it for q).

Since in accordance with (2.34)

$$\frac{u_{c\,\text{вых}}(T)}{U_{\text{ш}}} = \sqrt{\frac{2Q}{N_0}},$$

then

$$U_{\text{ш}} = \frac{u_{c\,\text{вых}}(T)}{\sqrt{2q}},$$

and formula (5.34b) may be represented also in the following form:

$$U_{\text{пор}} = \frac{u_{c\,\text{вых}}(T)}{\sqrt{q}} \sqrt{\ln \frac{1}{P_{\text{лт}}}}, \tag{5.34c}$$

where $u_{c\,\text{вых}}(T)$ -- maximum (peak) signal voltage value at correlator (or its replacement matched filter) output.

It follows from the formulas presented that, during binary signal detection, detection reliability (error probability) will not depend on signal $u_c(t)$ shape and will depend only on its energy Q.

5.3 Discrimination of Two Non-Zero Signals

In the preceding section, we examined a case where one of two possible signals [$u_0(t)$ and $u_1(t)$] is identical with zero. It is possible analogously to examine a more common case where both possible signals, $u_1(t)$ and $u_2(t)$, are non-zero signals. V. A. Kotel'nikov obtained the following results for this case [1]. Composite error probability $P_{\text{ош}}$ equals

$$P_{\text{ош}} = P(u_1) P(u_2 \text{ вм. } u_1) + P(u_2) P(u_1 \text{ вм. } u_2), \tag{5.35}$$

where $P(u_1)$ and $P(u_2)$ -- a priori probabilities of the presence of signals u_1 and u_2, while $P(u_2 \text{ вм. } u_1)$ and $P(u_1 \text{ вм. } u_2)$ -- conditional probabilities of reception of signal u_2 instead of u_1 and signal u_1 instead of u_2, respectively. Here /80

$$P(u_2 \text{ вм. } u_1) = V(\alpha_{21}), \tag{5.36}$$

where

$$V(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-z^2/2} \, dz; \tag{5.28a}$$

$$\alpha_{21} = \alpha + \frac{1}{2\alpha} \ln \frac{P(u_1)}{P(u_2)}; \tag{5.37}$$

$$\alpha^2 = \frac{1}{2N_0} \int_0^T [u_1(t) - u_2(t)]^2 \, dt. \tag{5.38}$$

It is sufficient in formulas (5.36)--(5.38) to replace indices 1 and 2 here and there to find probability $P(u_1 \text{ вм. } u_2)$.

Analysis of these relationships demonstrates that error probabilities $P(u_2 \text{ вм. } u_1)$, $P(u_1 \text{ вм. } u_2)$ and $P_{\text{ош}}$ decrease as magnitude α rises. But, it follows from formula (5.38) that magnitude α is maximum if

$$u_2(t) = -u_1(t); \tag{5.39}$$

therefore, relationship (5.39) insures receipt of minimum error probabilities. Here, signals $u_1(t)$ and $U_2(t)$ have identical energy

where

$$\left.\begin{aligned} & Q_1 = Q_2 = Q, \\ & Q_1 = \int_0^T u_1^2(t) \, dt; \quad Q^2 = \int_0^T u_2^2(t) \, dt \end{aligned}\right\} \tag{5.40}$$

and

$$\alpha^2 = \frac{2Q}{N_0}. \tag{5.41}$$

If signals $u_1(t)$ and $u_2(t)$ have identical energy ($Q_2 = Q_1$), but satisfy an orthogonality condition rather than condition (5.39), i. e.,

$$\int_0^T u_1(t)\, u_2(t)\, dt = 0, \tag{5.42}$$

then, from (5.38) we obtain:

$$\alpha^2 = \frac{Q}{N_0}. \tag{5.43}$$

Finally, in the case of binary detection, when one of the signals (u_2, for example) is identical with zero (passive spacing), from (5.38) one obtains

$$\alpha^2 = \frac{Q}{2N_0}. \tag{5.44}$$

In phase-shift keying [PSK] telegraph communications, relationship (5.39) /81 occurs, while relationship (5.42) occurs during frequency-shift key [FSK] with sufficient frequency f_1 and f_2 separation; binary detection corresponds to telegraph transmission with passive spacing.

Obtaining identical signal reproduction reliability in all these cases (i. e., identical error probabilities) requires an identical parameter α value. But, it follows from comparison of formulas (5.41), (5.43), and (5.44) that obtaining an identical α value requires signal energy Q greater by a factor of 2 for FSK and greater by a factor of 4 during passive spacing than during PSK. Consequently, FSK provides a loss of required energy greater by a factor of 2, and, for transmission with passive spacing, greater by a factor of 4 than is the case for PSK.

It should be noted here, however, that, in active spacing systems (FSK and

PSK, for example), energy Q will be expended during transmission of any of the signals (u_1 or u_2), while, in systems with passive spacing, energy is not expended in the spacings. Therefore, if pulses and spacings, for example, have identical duration and a priori probabilities $P(u_1)$ and $P(u_2)$, respectively, then the average energy that will be expended during a large number of tests*, in the case of passive spacing equals

$$Q_{\text{ср пп}} = P(u_1)\, Q_{\text{пп}},$$

then, in active spacing systems, it equals

$$Q_{\text{ср ап}} = Q_{\text{ап}};$$

therefore

$$\frac{Q_{\text{ср пп}}}{Q_{\text{ср ап}}} = P(u_1)\frac{Q_{\text{пп}}}{Q_{\text{ап}}}.$$

It follows from this that, given $P(u_1) \ll 1$, systems with passive spacing are more advantageous than those with active spacing (from the standpoint of expended average energy $Q_{\text{ср}}$), in spite of the fact that $Q_{\text{пп}}$ is several times greater than $Q_{\text{ап}}$.

If $P(u_1) = 0.5$, i. e., pulses and spacings are equally probable, then from the point of view of average energy $Q_{\text{ср}}$, passive spacing systems and FSK systems are equivalent, while a PSK system in comparison provides an energy saving greater by a factor of 2.

It also is evident that PSK is better than FSK, both from the standpoint of each signal's energy Q and of average energy $Q_{\text{ср}}$.

*Here, each test will comprise determination of which signal (u_1 or u_2) will be received during time interval (0, T), while only one of these signals may exist simultaneously in this interval.

5.4 Discrimination of m Orthogonal Equiprobable Signals Having Identical Energy /82

Calculation of receiver sensitivity is very complicated in the general case of signals (or messages) with many discrete values. Therefore, V. A. Kotel'nikov only computed for several frequent cases, the most significant being a case of discrimination of m orthogonal equiprobable signals having identical energy.

Here, it is assumed that the signal at receiver input may have one of m possible values $u_1(t)$, $u_2(t)$, . . ., $u_m(t)$.

All possible signals have identical energy

$$Q_1 = Q_2 = \ldots = Q_m = Q$$

and they satisfy the orthogonality condition, i. e.,

$$\int_0^T u_k(t)\, u_l(t)\, dt = 0 \quad \text{where } l \neq k \tag{5.45}$$

Since all signals have identical energy, then probability $P(u_0)$ of a zero signal equals zero. All non-zero signals are equally probable, i. e., the a priori probability of the presence of each signal equals

$$P(u_1) = P(u_2) = \ldots = P(u_m) = \frac{1}{m}. \tag{5.46}$$

The most-important individual instances of orthogonal signals are those which do not overlap in time (Figure 5.5) or in frequency spectrum (even though, generally speaking, those signals which overlap in time and in frequency spectrum, such as $\sin \frac{2\pi}{T} t$ and $\cos \frac{2\pi}{T} t$, also may be orthogonal).

The receiver's job is signal discrimination, i. e., determination of exactly which one of the m possible signals was present at receiver input during observation cycle (0, T). Here, interval T is selected in such a way that it encompasses all possible signal $u_k(t)$ positions.

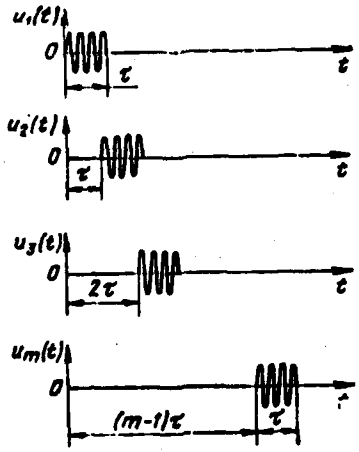


Figure 5.5

Thus, for example, if the possible orthogonal signals take the form depicted in Figure 5.5, then the following must be the case

$$T = m\tau, \tag{5.47}$$

where T -- duration of each signal.

The general optimum receiver schematic depicted in Figure 5.2 is valid for this case, but zero channel B_0 must be deleted.

Inverse probabilities $P_y(u_1) \ldots P_y(u_m)$ of all m possible signals /83 are compared in unit CC and that signal u_k, the inverse probability $P_y(u_k)$ of which turns out to be greatest, is selected.

Formula (5.4) is used in the general case to determine probability $P_y(u_k)$. In this instance, $P(u_k)$ and Q_k will not depend on k and formula (5.4) may be written in the following form:

where

$$\left.\begin{array}{l} P_y(u_k) = k_3 e^{\xi_k}, \\ \xi_k = \frac{2}{N_0}\int_0^T y(t)\, u_k(t)\, dt, \end{array}\right\} \quad (5.48)$$

k_3 — constant factor not depending on number k.

We will find correct signal discrimination probability $P_{\text{прав}}$. It is determined in the general case from formula (5.8). In the case examined, relationships (5.46) occur; in addition, due to the identical nature of all non-zero channels, this condition must be met

$$P_{u_1}(u_1) = P_{u_2}(u_2) = \ldots = P_{u_m}(u_m). \quad (5.49)$$

Therefore, formula (5.8) is simplified and takes the form

$$P_{\text{прав}} = P_{u_k}(u_k). \quad (5.50)$$

Here, $P_{u_k}(u_k)$ is the probability that, when signal u_k is present at input, the receiver will reproduce that very signal u_k, i. e., magnitude $P_y(u_k)$ will turn out to be the greatest.

Consequently, computation of probability $P_{\text{прав}}$ from formula (5.50) means assuming that signal $u_k(t)$ is at input, i. e.,

$$y(t) = u_k(t) + u_{\text{ш}}(t). \quad (5.51)$$

Substituting (5.51) into (5.48) and considering orthogonality condition (5.45), we will obtain

$$\left.\begin{aligned}
\xi_1 &= \frac{2}{N_0}\int_0^T u_m(t)\,u_1(t)\,dt;\\
\xi_2 &= \frac{2}{N_0}\int_0^T u_m(t)\,u_2(t)\,dt;\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
\xi_k &= \frac{2}{N_0}\int_0^T u_m(t)\,u_k(t)\,dt + \frac{2Q}{N_0};\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
\xi_m &= \frac{2}{N_0}\int_0^T u_m(t)\,u_m(t)\,dt.
\end{aligned}\right\} \qquad (5.52)$$

It was demonstrated in § 5.2 that magnitude $\frac{2}{N_0}\int_0^T u_m(t)u_k(t)dt$ has a normal /84 distribution with a zero average value and dispersion $\frac{2Q_k}{N_0}$ [see formulas (5.17 and (5.21)].

Consequently, in the case examined when signals have identical energy Q, all values of $\xi_1, \xi_2, \ldots, \xi_m$ (with the exception of ξ_k) determined from formulas (5.52) have the identical law of distribution with a zero average value and dispersion

$$\sigma_\xi^2 = \frac{2Q}{N_0}. \qquad (5.53)$$

Magnitude ξ_k also has a normal distribution with dispersion σ_ξ^2, but its average value equals not zero, but

$$\bar{\xi}_k = \frac{2Q}{N_0}. \qquad (5.54)$$

In addition, it may be demonstrated that, due to the orthogonality of signals $u_1(t), \ldots, u_m(t)$, random magnitudes $\xi_1, \xi_2, \ldots, \xi_m$ statistically are independent.

Correct reproduction will take place if probability $P_y(u_k)$ turns out to be the greatest, i. e., if ξ_k turns out to be the greatest of all values $\xi_1, \xi_2, \ldots, \xi_m$. This means that, if parameter ξ_k turns out to equal certain value z (regardless of which one), then all remaining (m—1) magnitudes $\xi_1, \xi_2, \ldots, \xi_m$ (except ξ_k) must be less than z.

Consequently, probability $P_{\text{прав}}$ is the probability that, if $\xi_k = z$ (where z may be any value), then all remaining (m—1) magnitudes $\xi_1, \xi_2, \ldots, \xi_m$ (except ξ_k) are less than z.

Since magnitudes $\xi_1, \ldots, \xi_m$ statistically are independent and have an identical law of distribution, then the probability that all (m—1) of such magnitudes simultaneously will turn out to be less than z equals

$$[P(\xi_l < z)]^{m-1}, \tag{5.55}$$

where $l \neq k$.

The probability that magnitude ξ_k will fall within the limits of z to z + dz equals

$$W_k(z)\,dz, \tag{5.56}$$

where $W_k(\xi_k)$ — magnitude ξ_k law of distribution.

Magnitude ξ_k is independent relative to all other values $\xi_1, \ldots, \xi_m$ (except ξ_k). Therefore, the probability that these relationships are satisfied simultaneously

$$\xi_l < z \quad (\text{where } l \neq k)$$

and

$$\xi_k = z - z + dz,$$

equals

$$dP = [P(\xi_l < z)]^{m-1} W_k(z)\,dz.$$

Since it is immaterial what value z is for correct reproduction, then /85

$$P_{\text{прав}} = \int_{-\infty}^{\infty} [P(\xi_l < z)]^{m-1} W_k(z)\,dz, \tag{5.57}$$

where $W_k(z)$ -- probability density of magnitude ξ_k, while $P(\xi_l < z)$ -- probability that $\xi_l < z$ (where $l \neq k$).

In the case examined, considering the laws of distribution of magnitudes $\xi_1, \xi_2, \ldots, \xi_k, \ldots, \xi_m$, we have

$$\left.\begin{aligned} W_k(z) &= \frac{1}{\sqrt{2\pi}\sigma_\xi} e^{-\left(z-\frac{2Q}{N_0}\right)^2 / 2\sigma_\xi^2}; \\ P(\xi_l < z) &= \frac{1}{\sqrt{2\pi}\sigma_\xi} \int_{-\infty}^{z} e^{-x^2/2\sigma_\xi^2}\,dx; \\ \sigma_\xi^2 &= \frac{2Q}{N_0}. \end{aligned}\right\} \tag{5.58}$$

Substituting (5.58) into (5.57), we obtain

$$1 - P_{\text{ош}} = P_{\text{прав}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[1 - V\left(\sqrt{\frac{2Q}{N_0}} + y\right)\right]^{m-1} e^{-y^2/2}\,dy,$$

where

$$V(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-t^2/2}\,dt. \tag{5.59}$$

Computations using this formula are possible only by means of numerical integration. However, if the error probability is slight ($P_{\text{ош}} \leqslant 0.1$), it is possible to simplify expression (5.59) considerably, having employed asymptotic expansion

of probability integral V(x). Here, (5.59) is reduced to an approximate formula [106]:

$$\frac{Q}{N_n} \approx 2 \ln \frac{1}{P_{\text{ош}}} + \ln (m-1) - 2{,}8. \qquad (5.60)$$

The error in this formula does not exceed 1 db if $P_{\text{ош}} \leqslant 0{,}1$.

5.5 The Case of an m-Channel Receiving Device

It is interesting to note that the relationships obtained for reception of one of m orthogonal equiprobable signals with identical energies are valid also for the m-channel receiving device whose block diagram is depicted in Figure 5.6.

This receiving device comprises m identical channels, at the input of which are active independent fixed noise voltages $u_{\text{ш}1}(t), u_{\text{ш}2}(t), \ldots, u_{\text{ш}m}(t)$ (internal /86

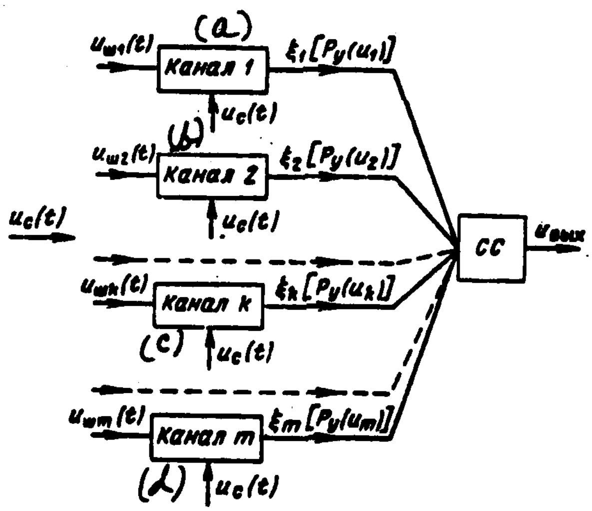


Figure 5.6. (a) -- Channel 1; (b) -- Channel 2; (c) -- Channel k; (d) -- Channel m.

noise of these channels, for example) with normal laws of distribution, zero average values, and identical dispersions N (where N = $\overline{u^2}_{\text{ш}k}$).

Signal $u_c(t)$, the shape of which is precisely known, may be present at input of one of these channels with equal probability (1/m). The job of the receiving

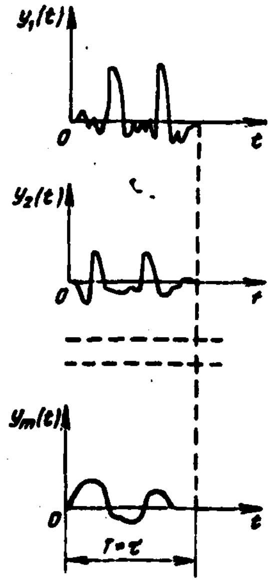


Figure 5.7

device is to determine the input of the exact channel at which a signal is present during a given observation cycle (0, T). The receiver has input voltage realizations $y_1(t)$, $y_2(t)$, . . ., $y_m(t)$ to solve this problem (Figure 5.7). It is known that one of these realizations will comprise the sum of noise $u_{mk}(t)$ and signal $u_c(t)$, while the remaining (m—1) realizations will comprise only noise. The task is to determine precisely which of these realizations will comprise the signal. Here, the assumption is that the signal will last the entire time

$$T = \tau.$$

Since noise is stationary, then it is possible to place realization $y_1(t)$, . . ., $y_m(t)$ in time, rather than in space, as depicted in Figure 5.8. Here, the task boils down to determining in which of m sectors of composite realization y(t) the signal with duration τ is located or, which is the same thing, determining which of m signals nonoverlapping in time (orthogonal) is in /87 this realization.

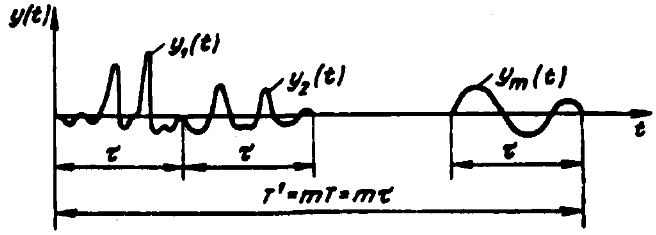


Figure 5.8

This problem was examined above and it was found that its solution requires comparison of inverse probabilities $P_y(u_k)$ or, which is the same thing, comparison of magnitudes ξ_k, and to select the greatest.

Therefore, in an m-channel receiver (Figure 5.6), for best solution of the problem, each channel also must comprise a computer of magnitude ξ_k, i. e., a correlator determining ξ_k from formula

$$\xi_k = \frac{2}{N_0} \int_0^T y_k(t)\, u_c(t)\, dt, \tag{5.61}$$

where $y_k(t)$ -- oscillation at input of the k-th channel. Comparison of magnitudes $\xi_1, \xi_2, \ldots, \xi_m$ and selection of the greatest one occurs in unit CC. As a result of this selection, the decision is made that the signal is present at the input of that channel in which magnitude ξ_k turned out to be greatest.

Since signal may be present in only one channel, the second channel, for instance, then $y_k(t)$ will comprise the signal-noise sum only in it:

$$y_2(t) = u_{m2}(t) + u_c(t);$$

in all other channels, it will be

$$y_k(t) = u_{mk}(t) \qquad \text{(where } k \neq 2)$$

It is not difficult to become convinced that, here, magnitudes $\xi_1, \xi_2, \ldots, \xi_m$ have exactly the same form as in the case of m orthogonal signals

[see expression (5.52)]. Therefore, formulas (5.57), (5.59), and (5.60) remain valid if, in these formulas, m is understood to mean the number of receiver channels, while probability $P_{\text{прав}}$ is understood to mean the probability of correct designation of the channel in which the signal is present.

CHAPTER SIX /88

RECEPTION OF INDIVIDUAL ANALOG MESSAGE VALUES

6.1 General Relationships

It is assumed during reception of individual analog message values that message x is an analog random magnitude, which is constant during an observation cycle (0, T), and changes from one interval to another with respect to random laws with a priori probability distribution P(x). The limits of possible message changes, $x_{мин}$ and $x_{макс}$, are assumed to be known.

Kotel'nikov for convenience assumes that

$$x_{мин} = -1;\ x_{макс} = 1. \quad (6.1)$$

Here, in accordance with a normality condition, the following must be true

$$\int_{-1}^{1} P(x)\,dx = 1. \quad (6.2)$$

A receiver computes inverse probability distribution $P_y(x)$ determined from formula (4.11) and supplies that message value $x_{ун}$ at which function $P_y(x)$ is maximum.

For simplicity, it is assumed that all message values are equiprobable, i. e.,

$$P(x) = \text{const}. \tag{6.3}$$

Here, expression (4.11) takes the form

$$P_y(x) = k_2 \exp\left[-\frac{1}{N_0}\int_0^T [y(t) - u_x(t)]^2\, dt\right]. \tag{6.4}$$

where constant k_2 will be found from normality condition (6.2).

The assumption that all message values are equally probable considerably simplifies the analysis and is proved, in addition, by the following circumstances:

1. In several cases, all message values actually are equiprobable.

2. Often there is absolutely no information available on the law of distribution P(x). Here, the assumption concerning equal probability of all messages is usually most natural.

In a number of cases, equidimensional distribution P(x) is the most favorable, i. e., requires the greatest signal energy during reception (see Chapter 17 for more details on this). Consequently, assuming distribution P(x) to be equidimensional, we thereby are investigating the worst case.

3. Given the requirement for highly-precise message reproduction, i. e., given a high signal-to-noise ratio, the structure of the optimum receiver /89 and its sensitivity essentially will not depend on the type of a priori distribution P(x) (proof of this point is provided in Chapter 19). Therefore, given a high signal-to-noise ratio, it is completely permissible for simplified analysis to assume that distribution P(x) is equidimensional, even if it is known that this distribution may be significantly irregular.

Since Kotel'nikov succeeded in conducting the most-detailed analysis of analog

message reception only for great signal-to-noise ratio values, then his assumption (6.3) is completely valid.

It follows from (6.4) that most-probable message value $x_{y\,n}$ must satisfy the equation*

$$\left[\frac{\partial \xi(x)}{\partial x}\right]_{x_{y\,n}} = 0, \tag{6.5}$$

where

$$\xi(x) = \frac{1}{T}\int_0^T [y(t) - u_x(t)]^2 dt = \langle [y(t) - u_x(t)]^2 \rangle_T. \tag{6.6}$$

In expanded form, equation (6.5) has the form

$$\left([y(t) - u_{x_{y\,n}}(t)] \left[\frac{\partial u_x(t)}{\partial x}\right]_{x_{y\,n}} \right)_T = 0. \tag{6.7}$$

Next, the form of function $P_y(x)$ approximating its maximum will be found, i. e., approximating most-probable value $x_{y\,n}$.

For x approximating $x_{y\,n}$, it is possible to assume

$$u_x(t) = u_{x_{y\,n}}(t) + (x - x_{y\,n}) \left[\frac{\partial u_x(t)}{\partial x}\right]_{x_{y\,n}}. \tag{6.8}$$

Substituting this expression into formula (6.4) and considering relationship (6.7), we obtain

$$P_y(x) = k_3 e^{-h(x - x_{y\,n})^2}, \tag{6.9}$$

*It is assumed that function $\xi(x)$ is differentiated with respect to x.

where

$$b=\frac{T}{N_0}\left(\left[\frac{\partial u_x(t)}{\partial x}\right]^2_{x_{y\,n}}\right)_T, \qquad (6.10)$$

while k_3 -- constant not depending on x and determined from normality condition

$$\int_{-1}^{1} k_3 e^{-b(x-x_{y\,n})^2}\,dx=1. \qquad (6.11)$$

If noise intensity N_0 is sufficiently slight (i. e., the signal-to-noise ratio is great), then the indicator of power in expressions (6.4) and (6.9), outside the area in which equality (6.8) is valid, will become so large relative /90
to absolute magnitude that it is possible to disregard magnitude $P_y(x)$ outside this area. Here, one may consider that distribution (6.9) is valid in the area of all possible values of x (and not only approximating $x_{y\,n}$) and, in expression (6.11), it is possible to change the integration limits by $-\infty$ and ∞. Then, from formula (6.11), we obtain

$$k_3=\sqrt{\frac{b}{\pi}}.$$

Consequently, given a high signal-to-noise ratio, distribution $P_y(x)$ has the form of a gaussian curve

$$P_y(x)=\sqrt{\frac{b}{\pi}}\,e^{-b(x-x_{y\,n})^2}. \qquad (6.12)$$

Signal reproduction quality is characterized by error

$$\delta=y-x=x_{y\,n}-x. \qquad (6.13)$$

Since x here is understood to mean a message normalized in such a way that its maximum magnitude equals unity, then error δ also turns out to be normalized accordingly.

The mean square of the error equals

$$\overline{\delta^2} = \overline{(x_{y\,н} - x)^2} = M(x_{y\,н} - x)^2. \qquad (6.14)$$

Two random magnitudes, x and $x_{y\,н}$, are included in this expression; here, magnitude $x_{y\,н}$ is random because it will depend on the type of realization y(t) of the signal-noise sum (see Figure 4.2).

Initially, we will find the conditional mean square of the error, determined by relationship

$$\overline{\delta_y^2} = M_y(x_{y\,н} - x)^2 = \int_{-1}^{1} (x_{y\,н} - x)^2 P_y(x)\,dx. \qquad (6.15)$$

It is evident from (6.15) that $\overline{\delta_y^2}$ will be found for given realization y, i. e., is conditional expected value M_y of the square of the error.

Introduction of magnitude $\overline{\delta_y^2}$ is convenient because its determination requires only knowing conditional distribution $P_y(x)$, which already had been found above.

For computation of unconditional mean square $\overline{\delta^2}$, magnitude $\overline{\delta_y^2}$ must, in turn, be averaged using multidimensional distribution P(y) of random function y(t). This second averaging in the general case encounters great difficulties. However, as Kotel'nikov demonstrated, given a sufficiently-great signal-to-noise ratio, magnitude $\overline{\delta_y^2}$ essentially will not depend on y and, consequently, it is possible to assume

$$\overline{\delta^2} = \overline{\delta_y^2}. \qquad (6.16)$$

Actually, for slight noise, from (6.12) and (6.15), we obtain /91

$$\overline{\delta_y^2} = \frac{1}{2b},$$

where b is determined from formula (6.10), i. e., it is a magnitude not depending

on $x_{y\,n}$, and, consequently, on y. Therefore, given slight noise, it is possible to assume

$$\overline{\delta^2}=\frac{1}{2b}. \tag{6.17}$$

It follows from formulas (6.10) and (6.17) that mean square error decreases with a rise in magnitude $\left(\left[\frac{\partial u_x(t)}{\partial x}\right]^2_{x_{y\,n}}\right)_T$, i. e., in the mean (for time T) square of the partial derivative of signal $u_x(t)$ with respect to message x. This result is fully understandable since the greater this derivative, the greater the change in oscillations $u_x(t)$ shape caused by slight change Δx in message x. Therefore, the greater the noise magnitude must be for the receiver to err in detecting this true message change.

Consequently, the greater magnitude $\left(\left[\frac{\partial u_x(t)}{\partial x}\right]^2_{x_{y\,n}}\right)_T$, the higher the receiver noise immunity.

It follows from relationships (6.12) and (6.13) that, given slight noise,

$$P_y(\delta)=\sqrt{\frac{b}{\pi}}e^{-b\delta^2}.$$

Since magnitude b will not depend on y in this case, then we obtain

$$P(\delta)=P_y(\delta)=\sqrt{\frac{b}{\pi}}e^{-b\delta^2}=\frac{1}{\sqrt{2\pi\overline{\delta^2}}}e^{-\delta^2/2\overline{\delta^2}} \tag{6.18}$$

i. e., error δ has a normal law of distribution P(δ) with a dispersion equalling $\overline{\delta^2}$.

Therefore, given slight noise, the probability that error δ will exceed magnitude ϵ with respect to absolute value equals

$$P(|\delta|>\varepsilon)=2V\left(\frac{\varepsilon}{\sqrt{\overline{\delta^2}}}\right), \tag{6.19}$$

where V(x) — probability integral determined from formula (5.28a).

The next important conclusion that Kotel'nikov obtained for a case of slight noise is that a receiver operating on the maximum inverse probability principle provides minimum possible mean square error. Proof of this assumption will not be provided here since it may be obtained as a partial case of more general theorems formulated in Chapter 19.

Thus, Kotel'nikov obtained the following important results for slight /92 noise, i. e., for a large signal-to-noise ratio.

1. A receiver operating on the maximum inverse probability density principle provides the minimally-possible mean square error. This error $\sqrt{\overline{\delta^2}}$ is determined from formulas (6.17) and (6.10).

2. Error δ is subordinate to a normal law of distribution with a zero average value and dispersion $\overline{\delta^2}$ [formula (6.18)].

Kotel'nikov succeeded in finding only the lower limit of the probability error for a random signal-to-noise ratio, determined by the following inequality,

$$P(|\delta|>\varepsilon)\geqslant\int_{-(1-\varepsilon)}^{1-\varepsilon}V(\alpha_1)\,dx_0, \tag{6.20}$$

where

$$\alpha_1^2=\frac{1}{2N_0}\int_0^T[u(x_0+\varepsilon,t)-u(x_0-\varepsilon,t)]^2\,dt;$$

$u(x_0\pm\varepsilon,t)$ -- function $u_x(t)$ values where $x=x_0\pm\varepsilon$.

In a number of cases, α_1 will not depend on x_0. Here, formula (6.20) is simplified and takes the form

$$P(|\delta|>\varepsilon)\geqslant 2(1-\varepsilon)V(\alpha_1). \tag{6.20a}$$

Based on the resultant general relationships, Kotel'nikov compared the various

types of modulation and reception methods. The main results of this analysis will be presented in the next section.

6.2 Comparison of Different Modulation Types and Reception Methods

In the case of amplitude modulation (AM)

$$u_x(t) = (1 + mx)\, B(t) \tag{6.21}$$

and

$$\frac{\partial u_x(t)}{\partial x} = mB(t).$$

Therefore, formula (6.10) takes the form

$$b = \frac{m^2}{N_0} \int_0^T B^2(t)\, dt = \frac{m^2 Q_0}{N_0}, \tag{6.22}$$

where $Q_0 = \int_0^T B^2(t)dt$ is the energy of carrier oscillation B(t), i. e., signal energy in the absence of modulation.

Substituting relationship (6.22) into formula (6.17), we obtain /93

$$\overline{\delta^2} = \frac{N_0}{2m^2 Q_0}. \tag{6.23}$$

In the case of frequency modulation (ChM) [FM]

$$u_x(t) = \sqrt{2} U_0 \cos[(\omega_0 + \Omega x)\, t + \varphi_0]. \tag{6.24}$$

Given slight noise, the mean square of error $\overline{\delta^2}$ is determined from formulas (6.10) and (6.17), i. e.,

$$\overline{\delta^2} = \frac{N_0}{2T\left(\left[\frac{\partial u_x(t)}{\partial x}\right]^2\right)_T}. \tag{6.25a}$$

where

$$\left(\left[\frac{\partial u_x(t)}{\partial x}\right]^2\right)_T = \frac{1}{T}\int_{t_1}^{t_1+T}\left[\frac{\partial u_x(t)}{\partial x}\right]^2 dt. \tag{6.25b}$$

For several signal types, the result of averaging using formula (6.25b) will not depend on moment t_1, when the observation began and it is possible to select any magnitude t_1, to consider, for example, that $t_1 = 0$ (as was the case above) or to assume that $t_1 = -T/2$. However, in a case of FM, magnitude t_1 significantly impacts upon averaging results.

Actually, from relationships (6.24), (6.25a), and (6.25b) (where $\omega T \gg 1$), we obtain

$$\overline{\delta^2} = \frac{3N_0}{2Q\Omega^2(T^2 + 3t_1T + 3t_1^2)}. \tag{6.26a}$$

This expression where $t_1 = -T/2$ has a maximum equalling

$$\overline{\delta^2} = \frac{6N_0}{Q\Omega^2T^2}. \tag{6.26b}$$

Where $t_1 > -T/2$, error $\overline{\delta^2}$ decreases monotonically when t_1 increases.

Consequently, removing moment t_1 when observation begins, it is possible <u>in principle</u> to make a message reproduction error as slight as desired. This result is explained as follows.

The following signals correspond to two message x values x_1 and x_2

$$u_{x1}(t) = U_0\sqrt{2}\cos[(\omega_0 + \Omega x_1)t + \varphi_0]$$

and

$$u_{x2}(t)=U_0\sqrt{2}\cos[(\omega_0+\Omega x_2)t+\varphi_0].$$

The difference in their phases equals

$$\Delta\varphi=\Omega(x_2-x_1)t=\Delta x\Omega t, \quad (6.26c)$$

where $\Delta x = x_2 - x_1$.

Consequently, signals $u_{x1}(t)$ and $u_{x2}(t)$ differ, not only with respect /94
to frequency, but also with respect to rf occupation phase. Here, it is evident from expression (6.26c) that if one begins observation of oscillations $u_{x1}(t)$ and $U_{x2}(t)$ over sufficiently-great time t following a frequency change, then even slight message change Δx will cause a very great signal phase change.

Consequently, signal phase modulation [PM] and FM occur, which makes it possible clearly to detect even the slightest message x changes if selected moment t, when the observation begins, is sufficiently late.

However, realization of very great noise immunity occurring during great t_1 changes is hindered by the following circumstances:

1. Since increased noise immunity is achieved due to phase modulation, initial signal rf occupation phase φ_0 must be precisely known at the point of reception, which does not occur in a majority of cases.

2. The lag in receiver message reproduction rises with a t_1 increase.

3. Formula (6.26a) is valid only given a great signal-to-noise ratio and, as demonstrated in the Kotel'nikov analysis, the greater the t_1, the greater the signal-to-noise ratio must be in order for the advantage provided by formula (6.26a) actually to occur.

It is shown in § 13.3 that, in those cases when phase φ_0 is unknown at the point of reception, error δ^T is determined by formula (6.26b). Therefore,

this case precisely is of the greatest practical value and, in future, we will assume that, given signal FM, message reproduction error is determined by formula (6.26b).

It follows from comparing formulas (6.26b) and (6.23) that FM provides a significant advantage compared to AM only where

$$\Omega T \gg 1. \tag{6.27}$$

But, where $\Omega T \gg 1$, the frequency band occupied by possible (anticipated) signals turns out to be considerably broader in the case of FM than in the case of AM.

Actually, let the carrier oscillation during the entire observation cycle have the form of a sinusoid with frequency f_0 (Figure 6.1a). The spectrum of

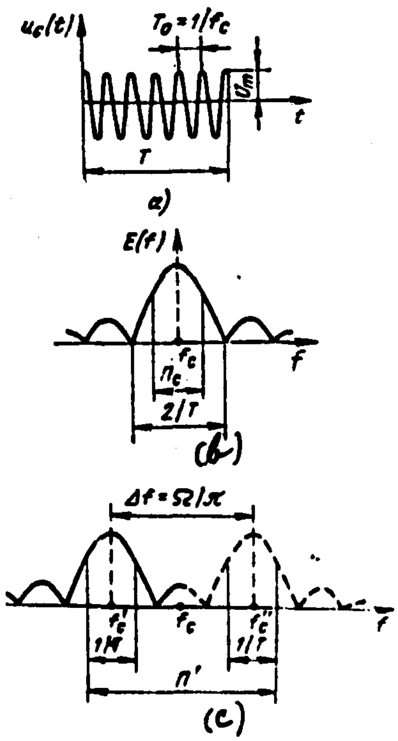


Figure 6.1

such an oscillation is depicted in Figure 6.1b. The basic part of the spectrum has width Π_c, equalling

$$\Pi_c \approx \frac{1}{T} \tag{6.28}$$

If such a signal is amplitude modulated by message x, then the shape of the spectrum remains unchanged for all values of x. Only the scale with respect to the Y-axis of curve E(f) changes [since, during interval (0, T), message x /95 is constant and changes only from one observation cycle to another]. Therefore, optimum receiver bandwidth must approximately equal Π_c :

$$\Pi \approx \Pi_c \approx \frac{1}{T}. \tag{6.29}$$

In the case of FM, the sinusoid frequency change ranges from $f_c' = f_c - \frac{\Omega}{2\pi}$ to $f_c'' = f_c + \frac{\Omega}{2\pi}$. Therefore, given extreme message x values (x = -1 and x = 1), anticipated signal spectra have the form depicted in Figure 6.1c and the optimum receiver must have band

$$\Pi' \approx \frac{1}{T} + \frac{\Omega}{\pi} \tag{6.30}$$

It follows from (6.29) and (6.30) that

$$\frac{\Pi'}{\Pi} \approx 1 + \frac{\Omega T}{\pi},$$

and, where $\Omega T \gg 1$, we obtain

$$\frac{\Pi'}{\Pi} \approx \frac{\Omega T}{\pi} \gg 1. \tag{6.31}$$

Thus, the advantage obtained during FM compared with AM results only due to receiver bandwidth expansion. But, noise intensity increases when the bandwidth is expanded and formula (6.26b) no longer is valid for great noise.

Kotel'nikov's approximate analysis of a case of great noise using inequality (6.20a) demonstrated that, given great noise, the advantage obtained with FM over AM decreases, while the decrease turns out to be more significant, the greater the ΩT. Consequently, the nature of the advantages FM has over AM, in the case of optimum receivers, is identical to that of known extant receivers; FM makes it possible to increase noise immunity only through expansion of receiver bandwidth and this may be realized only given a sufficiently-great signal-to-noise ratio, i. e., only when high message reproduction requirements are levied.

In the case of <u>pulse-position modulation</u> (VIM) [PPM], the high-frequency pulse envelope displaces in time, depending upon transmitted message x, without /96 changing its shape, i. e.,

$$u_x(t) = U_m\left(t - \frac{\tau_0}{2}x\right)\cos(\omega_0 t + \varphi). \quad (6.32)$$

Since observation time T must encompass all possible pulse positions during changes of x from -1 to 1, magnitude τ_0 must meet condition (Figure 6.2)

$$\tau_0 + \tau_u = T, \quad (6.33)$$

where τ_u — pulse duration.

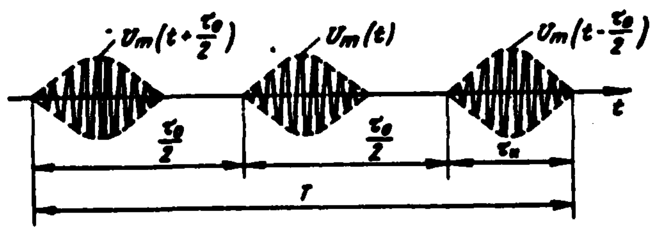


Figure 6.2

If

$$T \gg \tau_u,$$

then

$$\tau_0 \approx T \tag{6.33a}$$

The following formula is obtained from relationships (6.10), (6.17), (6.32), and (6.33a) for determination of mean square error for a great signal-to-noise ratio:

$$\overline{\delta^2} = \frac{4N_0}{\tau_0^2 T \left(\left[\frac{\partial U_m(t)}{\partial t}\right]^2\right)_T}. \tag{6.34}$$

It was assumed during derivation of this formula that envelope $U_m(t)$ is a slow time function compared with $\cos(\omega_0 t + \varphi)$.

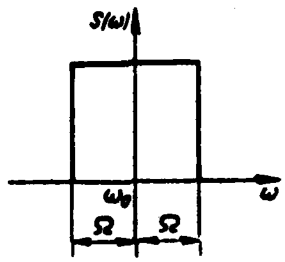


Figure 6.3

If the signal has a rectangular frequency spectrum (Figure 6.3), then its envelope is described by the equation

$$U_m(t) = U_0 \frac{\sin \Omega t}{\Omega t}. \tag{6.35}$$

Substituting this expression into formula (6.34) and assuming

$$\Omega T \gg 1, \tag{6.36}$$

we obtain

$$\overline{\delta^2} = \frac{12N_0}{\pi\tau_0^2 \Omega U_0^2}. \tag{6.37}$$

Signal relative energy Q in this case equals /97

$$Q = \frac{1}{2} T \langle U_m^2(t) \rangle_T \approx \frac{\pi}{2\Omega} U_0^2. \qquad (6.38)$$

Therefore, it is possible to write formula (6.37) in the following form:

$$\overline{\delta^2} = \frac{6N_0}{\Omega^2 \tau_0^2 Q}. \qquad (6.39)$$

Considering (6.33a), we obtain

$$\overline{\delta^2} \approx \frac{6N_0}{Q(\Omega T)^2}. \qquad (6.40)$$

Comparison of relationships (6.26b) and (6.40) demonstrates that, for the assumptions made, noise immunity is identical for PPM and FM. In a case of great noise, PPM and FM properties also approximately coincide: the advantage provided by PPM compared to AM (or pulse-amplitude modulation AIM [PAM]) decreases and is more significant, the greater the ΩT.

Comparison of the noise immunity of optimum and real receivers, given a high signal-to-noise ratio, will lead to the following results:

1. Given amplitude modulation, a receiver with a conventional amplitude detector and properly-selected bandwidths ahead of and beyond the detector provides noise immunity approximating the optimum.

2. Given pulse-position modulation, a real receiver also may provide noise immunity approximating the potential if the bandwidth selected in it is optimum and pulse envelope position with respect to time is computed with consideration of the bias both of the leading and of the trailing edge of this envelope [1].

6.3 Geometric Interpretation of Results

Results obtained become clearer when geometric models in n-dimensional space

are used. V. A. Kotel'nikov used n-dimensional geometry for the first time in the theory of signal reception [1], and it was used by a number of other authors, including A. A. Kharkevich [4].

Geometric interpretation is founded on use of expansion (1.12), which, in this case, it is convenient to reduce to the following form:

$$f(t) = \sum_{k=1}^{n} f_k' \psi_k'(t), \tag{6.41a}$$

where /98

$$f_k' = \frac{1}{\sqrt{2f_{\text{в}}}} f_k;\quad \psi_k'(t) = \sqrt{2f_{\text{в}}T}\,\psi_k(t). \tag{6.41b}$$

$$\left.\begin{aligned} &\int_0^T \psi_k'(t)\psi_l'(t)\,dt = 0 \ \text{with}\ l \neq k;\\ &\frac{1}{T}\int_0^T \psi_k'^2(t)\,dt = 1. \end{aligned}\right\} \tag{6.41c}$$

$$\sum_{k=1}^{n} f_k'^2 = \int_0^T f^2(t)\,dt = Q. \tag{6.41d}$$

Magnitudes f_k' and ψ_k' differ from f_k and ψ_k by constant factors and turn out to be more convenient because the mean square of magnitude ψ_k' equals unity, while the sum of the squares of f_k' ordinates equals oscillation $f(t)$ energy Q.

Functions $\psi_k'(t)$ are standard (known) unitary orthogonal time functions. Therefore, it follows from expression (6.41a) that any time function (with a restricted spectrum) completely is determined by n of its own values f_1', f_2', . . ., f_n', which may be considered coordinates of function f(t) in some imaginary n-dimensional space. Here, functions $\psi_k'(t)$ may be considered unitary directional orthogonal functions, i. e., unit vectors.

Space having more than three dimensions cannot be represented on a drawing, but it is possible to represent it qualitatively (conditionally) as depicted in

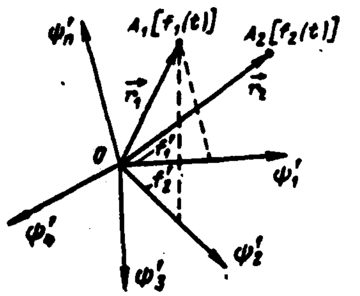


Figure 6.4

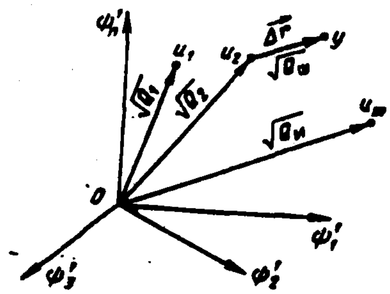


Figure 6.5

Figure 6.4. In the figure, ψ_1', ψ_2', . . ., ψ_n' are unitary vectors of the corresponding coordinate axes (unit vectors), f_1', f_2', . . ., f_n' are the coordinates of point A_1, i. e., ends of vector $\vec{r}_1$ or of the coordinates along the axes forming this vector.

Consequently, point A_1 (vector r_1) corresponds to some time function $f_1(t)$. Another time function, $f_2(t)$, has another combination of coordinates f_1', f_2', . . ., f_n' and, consequently, some other point A_2 (vector $\vec{r}_2$) of n-dimensional space corresponds to it.

The length of vector $\vec{r}$ in an orthogonal coordinate system is determined from the following expression:

$$r=\sqrt{f_1'^2+f_2'^2+\ldots+f_n'^2}=\sqrt{Q}, \tag{6.42}$$

i. e., the length of vector $\vec{r}$, which corresponds to oscillation f(t), equals the square root of the energy of this oscillation.

We will examine reception of discrete messages first. In this case, the signal may have one of m non-zero values $u_1(t)$, $u_2(t)$, . . ., $u_m(t)$ with energies

$Q_1, Q_2, \ldots, Q_m$. Some points $u_1, u_2, \ldots, u_m$ correspond to these possible signals in n-dimensional space (Figure 6.5).

For example, let signal $u_2(t)$ be present at receiver input during given /99 observation cycle (0, T). Due to presence of noise, total oscillation y(t) at receiver input equals

$$y(t) = u_2(t) + u_{\text{ш}}(t).$$

Since oscillation y(t) differs from $u_2(t)$ by the magnitude of noise voltage $u_{\text{ш}}(t)$, then corresponding point y in n-dimensional space (Figure 6.5) is separated from point u_2 by distance

$$\Delta r = \sqrt{Q_{\text{ш}}}, \tag{6.43}$$

where $Q_{\text{ш}}$ -- noise energy (conditional), i. e.,

$$Q_{\text{ш}} = \int_0^T u_{\text{ш}}^2(t)\,dt = \frac{1}{2f_{\text{в}}} \sum_{k=1}^{n} u_{\text{ш}k}^2. \tag{6.44}$$

Voltage $u_{\text{ш}}$ is a random time function. Therefore, vector $\overrightarrow{\Delta r}$ magnitude and direction also are random.

It follows from (6.44) that

$$\bar{Q}_{\text{ш}} = \frac{1}{2f_{\text{в}}} \sum_{k=1}^{n} \overline{u_{\text{ш}k}^2} = \frac{1}{2f_{\text{в}}} n\overline{u_{\text{ш}k}^2} = f_{\text{в}} T \frac{\overline{u_{\text{ш}k}^2}}{f_{\text{в}}}.$$

For white noise

$$\overline{u_{\text{ш}k}^2} = \frac{N}{f_{\text{в}}} = N_0$$

and

$$\bar{Q}_{\text{ш}} = N_0 f_{\text{в}} T. \tag{6.45}$$

Consequently, in the case of white noise

$$\Delta r_{эф} = \sqrt{\overline{\Delta r^2}} = \sqrt{\overline{Q_{ш}}} = \sqrt{N_0 f_э T}, \tag{6.46}$$

where $\Delta r_{эф}$ -- mean square value of distance Δr from total oscillation y(t) /100 to true signal $u_k(t)$ [in this case, $u_2(t)$].

It follows from formula (1.25) that, for white noise, the probability of a given realization $u_m(t)$ monotonically rises with a decrease in energy Q_m of this realization and has the greatest value where $Q_m = 0$.

Thus, in n-dimensional space, oscillation y(t) differs from true signal $u_k(t)$ by random vector $\overrightarrow{\Delta r}$. In the case of normal white noise, the most-probable value of distance Δr equals zero, while the mean square value is determined from formula (6.46). The greater distance Δr is, the less its probability.

The task of a receiver possessing oscillation y(t) is establishing exactly which of the m possible signals is present at input. A receiver operating on the maximum inverse probability principle each time selects that signal for which inverse probability $P_y(x)$ is maximum. If all signals are equally probable, then, as follows from (4.11), magnitude $P_y(x)$ is maximum if the following expression is minimum

$$\xi(x) = \int_0^T [y(t) - u_x(t)]^2 dt. \tag{6.47}$$

Consequently, given equiprobable signals, the receiver must each time select that signal $u_x(t)$ for which magnitude $\xi(x)$ is minimum. But, it follows from (6.47) that $\xi(x)$ is the energy (conditional) of the difference of oscillations y(t) and $u_x(t)$; the squares of distances in n-dimensional space correspond to energies. Therefore, given equiprobable signals, the optimum receiver each time must select the signal located closest in n-dimensional space to total oscillation y(t).

Such a result is fully understandable. Actually, it was demonstrated above that the shorter the distance Δr from oscillation y(t) to true signal $u_x(t)$, the

more probable it is. Therefore, if all anticipated signals are equally probable, that signal that is the least distance from oscillation y(t) should be selected as the true signal to obtain minimum error probability.

Now, we will examine reception of individual analog message values. In this case, message x may have any value ranging from -1 to 1; therefore, a continuous

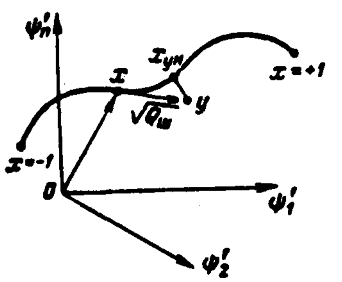


Figure 6.6

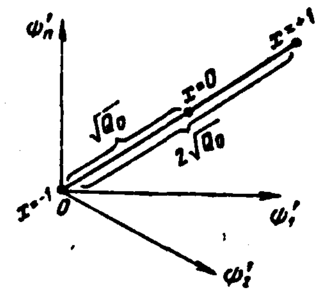


Figure 6.7

multidimensional line, called the signal line (Figure 6.6), rather than a collection of individual points, represents the aggregate of anticipated signals in n-dimensional space.

Point y, corresponding to total oscillation y(t), is separated from point x, corresponding to the signal carrying true message (x), by a distance equalling $\sqrt{Q_{ш}}$.

If, as Kotel'nikov assumed, all message x values are equally probable, then the receiver selects as the true value that message x value to which minimum magnitude $\xi(x)$ corresponds [formula (6.47)], i. e., that point on the signal line, /101
located the minimum distance from point y (point $x_{yн}$ in Figure 6.6). Here, a certain error is made

$$\delta = x_{yн} - x.$$

It is evident from Figure 6.6 that, for a given noise energy, the relative error will be less, the greater the entire signal line length. Therefore, one

should strive to increase signal line length to increase noise immunity.

For AM, only energy Q_x of oscillation $u_x(t)$ will depend on magnitude of x, while the shape of this oscillation remains unchanged. Therefore, the corresponding signal line has the form of an n-dimensional straight line, as depicted conditionally in Figure 6.7. For modulation capabability equalling 100%, signal energy changes from zero to $4Q_0$, where Q_0 -- carrier oscillation energy. It is possible here to increase signal line length only by increasing signal energy Q_0.

Consequently, for AM, it is possible to increase noise immunity (decrease mean square error $\overline{\delta^2}$) only by increasing energy Q_0 (for given noise intensity), which is fully compatible with relationship (6.23).

For FM, signal energy does not change in the modulation process. Consequently, all signal line points must be located from the origin at identical distance $\sqrt{Q}$, i. e., the signal line must be located on the surface of a hypersphere

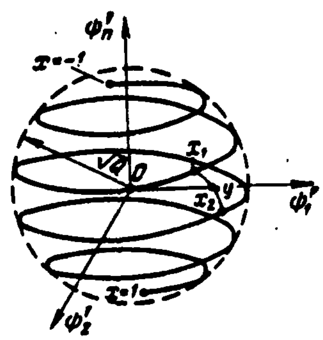


Figure 6.8

(in n-dimensional space, a hypersphere is the analog of a sphere of three-dimensional space). Therefore, in the case of FM, it is possible to draw the signal line as shown in Figure 6.8, i. e., in the form of some line on the surface of a sphere with a radius equalling $\sqrt{Q}$, where Q -- signal energy.

It is evident from comparing Figures 6.7 and 6.8 that, in the case of FM, signal line length may be much greater than for AM (given identical signal /102

energy) and, thanks to this, given slight noise, the mean square error will be significantly less. However, in the case of great noise, the advantage will disappear due to onset of so-called anomalous errors.

Actually, if noise energy $Q_ш$ is so great that point y is thrown aside from point x, corresponding to the true signal (and, thereby, to the true message), to a segment exceeding one-half the distance between adjacent turns of the spiral, then point x_2, located on the adjacent turn of the spiral, rather than point x_1 will be at the shortest distance from y. In this case, the receiver will reproduce message x_2, i. e., make an error corresponding to a complete turn of the spiral, rather than to a small portion of a turn. Such an abnormally-large error is called an anomalous error. It is evident that a threshold for onset of such an error exists--a slight signal-to-noise increase may lead to the transition from a normal error to an anomalous error.

For FM, an error decrease may be attained both through increasing signal energy Q and through increasing the number of dimensions n. Actually, the surface of the hypersphere rises with an increase in Q and, consequently, so does the length of the signal line located on it.

We will examine the simplest case, depicted in Figure 6.9, to explain the impact of the number of dimensions. Here, n equals unity, two, and three. It

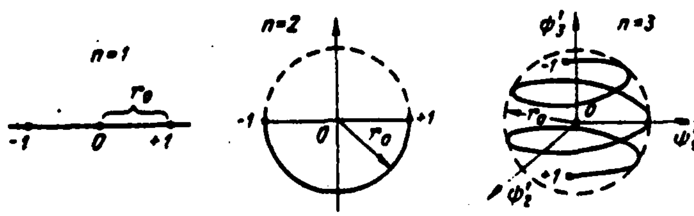


Figure 6.9

is evident from this figure that, given constant signal energy Q, i. e., given constant magnitude $r_0 = \sqrt{Q}$, possible signal line length also rises with a rise in n. But, n = $2f_вT$; consequently, in the case of FM, the error must decrease,

both with a rise in signal energy Q and with a rise in the number of dimensions, i. e., with a rise in the product of band $f_в$ during observation cycle T. This result fully is compatible with formula (6.26b). Mean noise energy [see formula (6.45)] also rises simultaneously with a rise in the number of dimensions $2f_вT$. Therefore, too great an increase in product $f_вT$ will lead to onset of anomalous errors and, consequently, to disappearance of the advantage FM supplies over AM.

Thus, all conclusions obtained in § 6.2 analytically have clear geometric confirmation.

In the case of PPM and other modulation methods in which oscillation energy remains unchanged in the modulation process, Figure 6.8 remains qualitatively valid and analogous results occur--where $f_вT \gg 1$, a significant advantage accrues compared with AM, but, here, a signal-to-noise ratio threshold value exists. /103
The probability of anomalous errors rises radically when this threshold is exceeded.

It is necessary to increase signal energy or the number of dimensions $2f_нT$, i. e., bandwidth or observation cycle, to decrease mean square error $\sqrt{\overline{\delta^2}}$ in the modulation types examined above.

Kotel'nikov examined the possibility of decreasing error $\overline{\delta^2}$, given constant signal energies and number of dimensions. If signal energy and number of dimensions remain unchanged, then the hypersurface on which the signal line is located also remains unchanged. Therefore, the signal line length increase required to decrease error may be achieved only by increasing the sinuosity of this line, i. e. for example, by more dense placement of the spiral turns. Evidently, great signal line sinuosity corresponds to a more complex type of signal modulation, such as double modulation. Therefore, when Q and $f_вT$ are unchanged, it is possible to achieve a decrease in error only through use of more complex modulation types. Since adjacent turns of the spiral (signal line) turn out here to be closer to each other, then anomalous errors will arise during a lesser signal-to-noise ratio than is the case for simple modulation.

Consequently, it is possible to reduce error when Q and $f_нT$ remain unchanged only in the event of . very slight signal-to-noise ratio. These qualitative results

stemming from representations in n-dimensional space are completely confirmed quantitatively.

Thus, for example, it has been demonstrated that if one simultaneously uses carrier FM and envelope PM with respect to the law

$$u_x(t) = U_0 [1 + \cos(\Omega_0 t + ax)] \cdot \cos(\omega_0 + \Omega x) t,$$

then the increase in parameter a makes it possible to decrease error $\overline{\delta^2}$ without a change in signal energy and spectrum width. However, the greater parameter a, given less signal-to-noise ratio, then anomalous errors arise and, consequently, the advantage obtained with such complex modulation disappears.

CHAPTER SEVEN /104

RECEPTION OF OSCILLATIONS

7.1 Basic Relationships

The term reception of oscillations is understood to mean reception of messages changing so rapidly that, during observation cycle (0, T), they must be considered time function x(t). It is assumed for simplicity that function x(t) is constrained by limits of ± 1.

The receiver computes inverse probability distribution $P_y(x)$ determined from formula (4.11) and supplies that message value $x_{yn}(t)$ at which function $P_y(x)$ is maximum:

$$\gamma(t)=x_{yn}(t). \qquad (7.1)$$

It is assumed that all possible realizations of the random function are equally probable, i. e.,

$$P(x)=\text{const} \quad \text{where} \quad |x|\leqslant 1;$$

$$P(x)=0 \quad \text{where} \quad |x|>1.$$

Here, the inverse probability density maximum corresponds to the minimum with respect to x of magnitude

$$\xi(x) = \int_0^T [y(t) - u_x(t)]^2\, dt. \tag{7.2}$$

Message reproduction error equals

$$\delta(t) = \gamma(t) - x(t) = x_{y\,\text{п}}(t) - x(t), \tag{7.3}$$

i. e.,

$$\gamma(t) = x(t) + \delta(t), \tag{7.4}$$

where $\gamma(t)$ -- oscillation at receiver output.

It is assumed that, when there is no noise, the receiver does not supply distortions, i. e.,

$$\gamma(t) = x(t);$$

consequently, function $\delta(t)$ also may be considered a component of output oscillation $\gamma(t)$ caused by noise action, i. e., a noise oscillation at receiver output. Here, since function x(t) is normalized so that it is dimensionless and included in the range of $\pm$ 1, then oscillations $\gamma(t)$ and $\delta(t)$ also are normalized.

The real (non-normalized) oscillation at receiver output may be written in the form

$$u_{\text{вых}}(t) = u_{\text{с}}(t) + u_{\text{шс}}(t), \tag{7.5}$$

where $u_{\text{с}}(t)$ -- usable voltage at receiver output obtained in the absence of /105 noise, while $u_{\text{шс}}(t)$ -- additional fluctuating component created by noise action.

Consequently,

$$u_{\Omega}(t) = kx(t), \tag{7.6}$$

where k -- some proportionality coefficient.

Let $U_{\text{м}\Omega}$ be the maximum value of output oscillation $u_{\Omega}(t)$, corresponding to $x(t) = x_{\text{макс}} = 1$; then, from (7.6) we have

$$k = U_{\text{м}\Omega}. \tag{7.7}$$

Substituting (7.6) and (7.7) into (7.5), we obtain

$$u_{\text{вых}}(t) = U_{\text{м}\Omega}\left[x(t) + \frac{u_{\text{ш}\Omega}(t)}{U_{\text{м}\Omega}}\right]. \tag{7.8}$$

It follows from comparison of expressions (7.4) and (7.8) that

$$\delta(t) = \frac{u_{\text{ш}\Omega}(t)}{U_{\text{м}\Omega}},$$

i. e.,

$$u_{\text{ш}\Omega}(t) = U_{\text{м}\Omega}\,\delta(t). \tag{7.9}$$

Kotel'nikov succeeded in concluding the mathematical analysis of oscillation reception, only given the following basic assumptions:

1. The signal-to-noise ratio is sufficiently slight.

2. Message x(t) will not comprise frequencies higher than some frequency limit $F_{\text{в}}$, which is much less than signal carrier frequency f_0.

3. Receiver structure is such that, when there is no noise, the receiver reproduces the message absolutely correctly; the structure remains unchanged when there is noise.

The concept of these assumptions will be examined in § 7.3.

The following basic results were obtained, given the aforementioned assumptions [1]:

1. Function $\delta(t)$ is a stationary fluctuating oscillation with a normal law of distribution with a zero average value. Therefore, $\delta(t)$ is characterized fully by its own power spectrum* $E^2_m(f)$ and

$$\overline{\delta^2} = \int_{F_n}^{F_B} E^2_m(f)\,df, \tag{7.10}$$

where F_n -- lowest frequency contained in message x(t).

2. A receiver operating on the maximum inverse probability density principle insures receipt of the minimum mean square error and minimum signal-to-noise ratio at receiver output.

3. The type of power spectrum $E_m^2(f)$ will depend on modulation type. /106 Possible modulation types are categorized as direct and indirect. Those modulation types in which signal $u_x(t)$ is linked with message x(t) with the aid of some type of operator (an integral or differentiator, for instance) are referred to as indirect modulation types. If message x(t) is included directly in signal $u_x(t)$, rather than under the sign of an operator, then the corresponding modulation type is referred to as direct modulation.

For AM

$$u_x(t) = U_0[1 + mx(t)]\cos(\omega_0 t + \varphi_0). \tag{7.11}$$

For PM

$$u_x(t) = U_0 \cos[\omega_0 t + m_\Phi x(t)]. \tag{7.12}$$

*It is assumed that spectrum $E_m^2(f)$ will not comprise negative frequencies.

For FM

$$u_x(t) = U_0 \cos\left[\omega_0 t + \Omega \int x(t)\, dt\right]. \tag{7.13}$$

Consequently, AM and PM are direct modulation types, while FM is an indirect type.

Those indirect modulation types in which x(t) is included under the sign of an integral are referred to as integral modulation types. Consequently, FM is an integral modulation type.

For direct modulation types we obtain

$$E_m^2(f) = \frac{N_0}{\left(\left[\frac{\partial u_x(t)}{\partial x}\right]^2\right)_T}, \tag{7.14}$$

and for integral types

$$E_m^2(f) = \frac{N_0 (2\pi f)^2}{\left(\left[\frac{\partial u_x(t)}{\partial \psi_x}\right]^2\right)_T}, \tag{7.15}$$

where $\psi_x = \int x(t)\, dt$.

The formulas have a slightly more complex form for pulse modulation types [1].

It follows from formulas (7.14) and (7.15) that the noise power spectrum at receiver output in the case of direct modulation types is equidimensional within a band range of $(F_x - F_n)$, while it is parabolic for integral modulation.

Substitution of expressions (7.11)--(7.13) into formulas (7.14) and (7.15) provide the following results:

for AM

$$E_m^2(f) = \frac{2N_0}{U_0^2 m^2}; \tag{7.16}$$

for PM

$$E_{ш}^{2}(f)=\frac{2N_0}{U_0^2 m_\phi^2}; \qquad (7.17)$$

for FM /107

$$E_{ш}^{2}(f)=\frac{2N_0}{U_0^2}\left(\frac{2\pi f}{\Omega}\right)^2. \qquad (7.18)$$

We will designate $U_{ш\Omega}$ -- mean square noise oscillation value at receiver output

$$U_{ш\Omega}=\sqrt{\overline{\overline{u_{ш\Omega}^{2}}}}. \qquad (7.19)$$

It follows from formulas (7.9), (7.10), and (7.19) that

$$\frac{U_{ш\Omega}^{2}}{U_{м\Omega}^{2}}=\int_{F_н}^{F_в} E_{ш}^{2}(f)\,df, \qquad (7.20)$$

where $U_{ш\Omega}/U_{м\Omega}$ -- ratio of the mean square noise oscillation value to maximum signal oscillation value.

Substituting expressions (7.16), (7.17), and (7.18) into formula (7.20) and assuming for simplicity that $F_н$ = 0, we obtain:

for AM

$$\frac{U_{ш\Omega}}{U_{м\Omega}}=\frac{\sqrt{2N_0F_в}}{U_0 m}; \qquad (7.21)$$

for PM

$$\frac{U_{ш\Omega}}{U_{м\Omega}}=\frac{\sqrt{2N_0F_в}}{U_0 m_\phi}; \qquad (7.22)$$

for FM

$$\frac{U_{\text{ш}\Omega}}{U_{\text{м}\Omega}} = \frac{\sqrt{2N_0F_{\text{в}}}}{U_0} \cdot \frac{2\pi F_{\text{в}}}{\sqrt{3}\,\Omega}; \tag{7.23}$$

for sinusoidal FM

$$\Omega = 2\pi\Delta f_m \tag{7.24}$$

and

$$\frac{U_{\text{ш}\Omega}}{U_{\text{м}\Omega}} = \frac{\sqrt{2N_0F_{\text{в}}}}{U_0} \cdot \frac{1}{\sqrt{3}\,m_{\text{ч}}}, \tag{7.25}$$

where

$$m_{\text{ч}} = \frac{\Delta f_m}{F_{\text{в}}} \tag{7.26}$$

is the modulation index, while Δf_m -- frequency deviation.

Formulas (7.21)--(7.23) coincide with the corresponding formulas for real AM, PM, and FM signal receivers.

Note 1. The term real FM (PM) signal receivers here and in future is understood to mean receivers comprising an rf amplifier, amplitude limiter, frequency (phase) detector, and af amplifier. The term real AM signal receiver is understood /108 to mean a receiver comprising an rf amplifier, amplitude (nonsynchronous) detector, and af amplifier. Formula (7.21) is valid for a real AM signal receiver with a synchronous detector not only given slight, but given essentially any, signal-to-noise ratio.

Note 2. Strictly speaking, relationships (7.21)--(7.25) occur for real receivers only in the event that, during computation of noise voltage $U_{\text{ш}\Omega}$, the signal is assumed to be unmodulated. Therefore, they are precise only given a slight signal modulation capability, i. e., given slight m, m_{Φ}, and $m_{\text{ч}}$. However,

it is possible to consider them approximately valid also for large m, m_ϕ , and $m_\text{ч}$ values.

Consequently, given a slight signal-to-noise ratio, the noise immunity of real AM, PM, and FM signal receivers coincides with potential noise immunity. Hence, it follows also that the advantages of PM and FM compared to AM in the case of optimum receivers are identical to those in the case of real receivers--given slight noise, use of PM and FM provides an advantage that is greater, the greater m_ϕ and $m_\text{ч}$, respectively. However, the greater m_ϕ and $m_\text{ч}$, then, given less noise, anomalous errors arise ameliorating this advantage.

Given PPM, based on the law provided in § 6.2 we obtain

$$E_m^2(f) = \frac{12N_0}{U_9^2}\left(\frac{F_\text{н}}{\Omega}\right)^2, \qquad (7.27)$$

where $F_\text{н}$ — pulse repetition frequency [PRF]; U_9 — effec' ive signal amplitude characterizing its average power equalling

$$U_9 = U_0\sqrt{\frac{\pi F_\text{н}}{2\Omega}}.$$

If the minimally-permissible PRF is selected, that being $F_\text{н} = 2F_\text{в}$, then formula (7.27) takes the form

$$E_m^2(f) = \frac{48N_0}{U_9^2}\left(\frac{F_\text{в}}{\Omega}\right)^2. \qquad (7.27a)$$

Considering expression (7.20) and assuming that $F_\text{н} = 0$, we obtain

$$\frac{U_{m\Omega}}{U_{\text{ш}\Omega}} = \frac{\sqrt{N_0F_\text{в}}}{U_9}\cdot\frac{4\sqrt{3}F_\text{в}}{\Omega}. \qquad (7.28)$$

Comparing formulas (7.23) and (7.28) and considering that, for FM

$$U_9 = \frac{U_0}{\sqrt{2}},$$

it is easy to become convinced that, given identical F_s and Ω and equal effective amplitudes U_s, and, consequently, equal average signal powers in the case of PPM, resultant ratio $U_{m\Omega}/U_{w\Omega}$ is greater by a factor of approximately 2 than is the case for FM. Consequently, potential noise immunity is somewhat better for FM than for PPM.

7.2 Impact of Signal Modulation Parameters /109

Kotel'nikov obtained all formulas presented in the previous section in the assumption that signal-to-noise ratio is sufficiently slight. Here, for the modulation types examined, the noise immunity of real receivers equals potential noise immunity or approximates it (given good limiters and proper selection of bandwidth ahead of and beyond the demodulator). Therefore, it is possible in the first approximation to assume that, if an optimum receiver must operate always in a slight signal-to-noise ratio mode, then its structure may be identical to that of real receivers, i. e., it may comprise an rf amplifier, demodulator, and af amplifier.*

Let the ratio of the effective noise and signal voltages at demodulator input equal U_{m}/U_{c} . Then, in order for the § 7.1 formulas to be valid, it is necessary to meet the condition

$$\frac{U_{m}}{U_{c}} \ll 1, \tag{7.29}$$

i. e., the signal-to-noise ratio must be slight already at demodulator input, not only at its output.**

*Here and in future, the term rf amplifier is understood to mean all linear stages ahead of the demodulator (including frequency converters as well in the case of a superhetrodyne receiver), while the term af amplifier is understood to mean all linear stages connected beyond the demodulator. It is assumed that the amplitude limiter (for FM and PM) and minimum limiter (in the case of pulse modulation) are part of the demodulator.

**In the case of an AM signal receiver with synchronous detector, as already noted in § 7.1, there is no requirement to meet condition (7.29).

Permissible signal-to-noise ratio $U_{ш\Omega}/U_{и\Omega}$ or $U_{ш\Omega}/U_{ф\Omega}$ at receiver output is set during receiver design since this is precisely what determines message reproduction quality. The higher the levied reproduction quality, the lower ratio $U_{ш\Omega}/U_{и\Omega}$ must be. However, given the possibility of optimum signal parameter selection, it turns out that condition (7.29) is not met for the majority of modulation types, even if requisite ratio $U_{ш\Omega}/U_{и\Omega}$ at output is very slight.

For example, we will examine the case of an FM signal.

If condition (7.29) is met, then formula (7.23) is valid. Thus, based on this, one should increase signal frequency deviation $\Omega/2\pi$ in order to improve noise immunity. But, both rf amplifier bandwidth P (where $m_ч \gg 1$, we obtain $П \approx 2\Omega/2\pi$) also correspondingly must be increased here, as a result of which noise voltage $U_ш$ rises. Until condition (7.29) is met, formula (7.23) is valid and, consequently, it is advantageous to continue to increase signal frequency deviation. A further parameter Ω increase may turn out to be inadvisible only if condition (7.29) is so far from being met that formula (7.23) validity is destroyed and a further Ω increase begins to lead in actuality to a ratio $U_{ш\Omega}/U_{и\Omega}$ increase, rather than its decrease. /110

Hence it follows that the optimum will be that signal frequency deviation value $\Omega_{опт}/2\pi$ at which condition (7.29) is not met and formula (7.23) a fortiori is imprecise, even if the signal-to-noise ratio at output must be slight. Thus, for example, analysis demonstrates that, in the case of sinusoidal FM where modulation index value $m_ч$ providing the minimum $U_{ш\Omega}/U_{и\Omega}$ ratio is optimum, this ratio is determined already not from formula (7.25), but from the following approximate relationship:

$$\frac{U_{ш\Omega}}{U_{и\Omega}} \approx \frac{\sqrt{2N_0F_в}}{U_0}\cdot\frac{1}{m_ч}, \tag{7.30}$$

i. e., formula (7.25) provides an error of almost a factor of 2. Consequently, formula (7.25) will be valid for FM only in the event, along with meeting the condition

$$\frac{U_{ш\Omega}}{U_{и\Omega}} \ll 1$$

signal modulation index $m_э$ will be significantly less than value $m_{э\,опт}$ for a given ratio $U_{m\Omega}/U_{м\Omega}$ (the less $U_{m\Omega}/U_{м\Omega}$, the greater $m_{э\,опт}$).

Analogously, for PM, one may demonstrate that some PM coefficient optimum value $m_{\Phi\,опт}$ exists, as does signal spectrum bandwidth value $2\Omega_{опт}/2\pi$ for PPM. Condition (7.29) is not met where the values of these parameters are optimum and the formulas presented in § 7.1 are imprecise for a slight signal-to-noise ratio, even given very slight $U_{m\Omega}/U_{м\Omega}$ ratios.

For AM, ratio $U_{m\Omega}/U_{м\Omega}$ will depend only on those signal parameters (U_0 and m), which do not impact upon its spectral width. Therefore, optimum signal parameter selection here does not cause a condition (7.29) disruption.

The analysis performed permits the following conclusions to be made:

1. Formulas (7.21)--(7.25) are sufficiently precise only when condition (7.29) is met, i. e., for a slight signal-to-noise ratio at demodulator input.*

2. Condition (7.29) is not met for all the modulation types examined above, with the exception of AM, given optimum signal parameters, even for a very slight $U_{m\Omega}/U_{м\Omega}$ ratio, and formulas (7.21)--(7.25) may provide considerable error. Therefore, in such a case, it is possible to use the aforementioned formulas only for a rough approximation.

3. It follows from point 2 that, for the majority of modulation types (PM, FM, PPM, and others), the formulas presented above are precise only for the /111 corresponding signal parameter values less than optimum, i. e., when the following conditions are met, respectively:

a) For FM

b) For PM

c) For PPM

$$\left.\begin{aligned} & m_э \ll m_{э\,опт}; \\ & m_\Phi \ll m_{\Phi\,опт}; \\ & \Omega \ll \Omega_{опт}. \end{aligned}\right\} \tag{7.31}$$

*See the note to condition (7.29).

7.3 Impact of Receiver Frequency Characteristic Shape

1. General Comments

We will examine the concept for assumption 3 made in § 7.1 relative to optimum receiver structure.

In accordance with this assumption, receiver structure for slight noise must be such that it will supply no message x(t) distortions when there is no noise, while the structure remains unchanged when noise appears.

In the general case, one may consider that, given noise, the resultant error $\delta_p(t)$ in message reproduction comprises two components:

$$\delta_p(t) = \delta(t) + \delta_д(t), \tag{7.32}$$

where $\delta_д(t)$ -- dynamic error, i. e., the error which would occur when there is no noise, while δ (t) -- additional error component caused by noise action and referred to in future as fluctuating error.

Error $\delta_д(t)$ was referred to as dynamic because it is caused primarily by the irregularity of receiver frequency characteristics (ahead of and beyond the demodulator), i. e., its inertness, and it rises with an increase in the rate of message x(t) change, i. e., its "dynamicity."

In accordance with the Kotel'nikov assumption, the structure of the optimum receiver, given slight noise, must be such that dynamic error $\delta_д(t)$ is identical with zero. However, this assumption is not always valid.

If noise intensity is infinitely slight, then it would be senseless to use that receiver structure in which dynamic error is not identical with zero since, in this case, the mean square of resultant error $\overline{\delta_p^2}$ would be a fortiori greater than for the receiver structure which insures $\delta_д(t) \equiv 0$. But, given infinitely-slight noise intensity, the task of insuring noise immunity becomes completely senseless. If noise intensity is only very slight, but finite, then it may /112 turn out to be advisible to narrow the bandwidth (ahead of and beyond the demodulator), even if some dynamic errors result.

Actually, narrowing the frequency band, causing a rise in dynamic error, simultaneously decreases the fluctuating error; therefore, the magnitude of the resultant error here may decrease. Consequently, even given very slight noise, it is necessary in the general case to strive to select receiver rf and af frequency characteristics (i. e., ahead of and beyond the demodulator) to insure minimum resultant error δ_p .

For simplicity, we will evaluate the errors based on their mean squares. Then, the optimum receiver will be the one insuring minimum mean square $\overline{\delta_p^2}$ of the resultant error.

Consequently, it is necessary to select receiver rf and af frequency characteristic shapes from the minimum mean square of the resultant error $\overline{\delta_p^2}$ in message x(t) reproduction.

2. Connection of Optimum Linear Filters

Initially, we will examine the methodology for finding the optimum shape of the receiver af frequency characteristic (i. e., beyond the demodulator).

Here, we will assume for simplicity that the rf bandwidth is so broad that no dynamic errors arise in the rf stages.

Also, we will assume that condition (7.29) is met. Then, the assumption of the possibility of dynamic errors in af stages is the only difference between

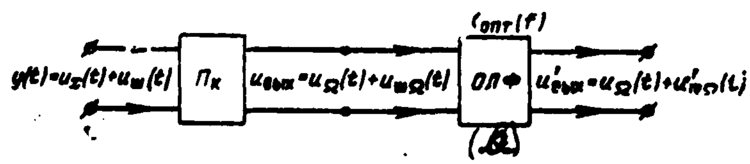


Figure 7.1. (a) -- Optimum linear filter (OLF).

the optimum receiver examined and the optimum Kotel'nikov receiver. Therefore, such a generalized optimum receiver must comprise receiver $П_к$, optimum in the

Kotel'nikov sense, and optimum linear filter (OLF) connected at its output (Figure 7.1).

The OLF must insure receipt of the minimum mean square error in reproduction of usable oscillation $u_\Omega(t)$, distorted by additive noise $u_{ш\Omega}(t)$, arriving at its input. Such a filter was examined in §2.2 for a case when $u_\Omega(t)$ is a stationary random process. For simplicity, we will not burden this filter with the condition of physical realization. In addition, we will assume that usable message /113
$u_\Omega(t)$ and fluctuating error component $u_{ш\Omega}(t)$ statistically are independent [the validity of such an assumption is greater, the less the signal is capable of being modulated by message x(t)]. Then, the filter frequency characteristic must meet condition (2.19), which may be written in the following form:

$$K_{опт}(f) = \frac{1}{1 + E^2_ш(f)/E^2_x(f)}, \qquad (7.33)$$

where $E^2_x(f)$ and $E^2_ш(f)$ -- power spectra of oscillations x(t) and δ(t), respectively, at Kotel'nikov receiver output. Here, it is assumed that message x(t) may be considered a stationary random process.

Since

$$u_\Omega(t) = U_{м\Omega}\, x(t) \quad \text{and} \quad u_{ш\Omega}(t) = U_{м\Omega}\, \delta(t),$$

then, the power spectra of oscillations $u_\Omega(t)$ and $u_{ш\Omega}(t)$ equal $U^2_{м\Omega}E^2_x(f)$ and $U^2_{м\Omega}\, E^2_ш(f)$, respectively. Therefore, formula (2.20) for the mean square error at filter output may be written in the following form:

$$\overline{\varepsilon^2} = U^2_{м\Omega} \int_0^\infty \frac{E^2_ш(f)}{1 + E^2_ш(f)/E^2_x(f)}\, df, \qquad (7.34)$$

where

$$\varepsilon(t) = u'_{вых}(t) - u_\Omega(t) = u'_{ш\Omega}(t). \qquad (7.35)$$

It is accepted in formula (7.34), as opposed to (2.20), that spectra $E_x^2(f)$ and $E_ш^2(f)$ will not comprise negative frequencies.

It follows from formulas (7.33)--(7.35) that

$$\frac{U'_{ш\Omega}}{U_{м\Omega}} = \sqrt{\int_0^\infty E_ш^2(f) K_{опт}(f)\, df}\,, \tag{7.36}$$

where

$$U'_{ш\Omega} = \sqrt{\overline{\varepsilon^2}} = \sqrt{\overline{u'^2_{ш\Omega}}}\,. \tag{7.37}$$

Formula (7.36) determines the relative mean square error at a generalized optimum receiver output. One may make the following conclusions from comparison of formula (7.36) with the corresponding formula (7.20) obtained for the optimum Kotel'nikov receiver:

1. There is no requirement to know the high and low frequency of the message x(t) spectrum for a generalized optimum receiver as is the case for the optimum Kotel'nikov receiver. This is more convenient since, in many actual cases, message spectrum $E_x^2(f)$ does not have sharply-defined cutoff in the region of the /114 high and low frequencies and it is unclear what $F_н$ and $F_в$ in formula (7.20) should be understood to mean.

2. Even if frequencies $F_н$ and $F_в$ are precisely known, one obtains

$$\frac{U'_{ш\Omega}}{U_{м\Omega}} \leqslant \frac{U_{ш\Omega}}{U_{м\Omega}}, \tag{7.38}$$

i. e., connection of the optimum linear filter at receiver output may provide a decrease in relative message reproduction error. This decrease is all the more significant, the greater the signal-to-noise ratio and the longer the curve $E_x^2(f)$ "tails" in the region of high and low frequencies.

Examples illustrating these postulations are presented in [125].

PART THREE /115

OPTIMAL RECEPTION OF SIGNALS WITH RANDOM PARAMETERS (ANALYSIS USING THE INVERSE PROBABILITIES APPROACH)

CHAPTER EIGHT

GENERAL RELATIONSHIPS

8.1 Problem Formulation

In the theory presented in Part II, Kotel'nikov assumes that signal $u_x(t)$ is precisely known, while the best receiver is the one which selects the message x value with maximum inverse probability $P_y(x)$ as the true value. Therefore, the Kotel'nikov approach may be called the maximum inverse probability approach, applicable to precisely-known signals.

Following publication of Kotel'nikov's work in 1946 [1], the theory of optimum reception methods developed, first, through examination of signals with random parameters and, second, through generalization of receiver optimization. The first important results along these lines were obtained in the works of Woodward and Davis [2, 28, and others].

As demonstrated in § 5.1, for given signal-plus-noise y(t), maximum usable

information on message x achievable at the point of reception is inverse probability distribution $P_y(x)$. Therefore, the job of the optimum receiver is to compute distribution $P_y(x)$. A decision on what message x was must be made on the basis of analysis of this distribution type.

A decision may be made in the simplest case from the maximum inverse probability principle that Kotel'nikov used. However, in several instances, the maximum inverse probability principle is not best, as even Kotel'nikov demonstrated. Therefore, in the general case, it is possible to confirm only that the optimum receiver must decide (using some specific method) based on analysis of inverse probability distribution $P_y(x)$. This general principle of finding the optimum decision will be called the inverse probability approach. Then, it is possible to consider the maximum inverse probability criterion one specific variant of the overall approach.

Examples of using the inverse probability approach to solve various /116
problems of receiving signals with random parameters in the presence of additive random noise $u_{\text{ш}}(t)$ are examined in subsequent chapters (9-13). Basic analysis will be performed assuming that this is normal white noise. However, in § 12.9, results obtained are generalized for random noise with an irregular power spectrum.

As was the case in Part II, analysis initially will be performed for discrete messages, then we will look into receipt of individual analog message values and reception of oscillations. Here, considerably more attention is devoted to different signal detection cases, i. e., actually establishing whether or not a signal is present at receiver input. This is because, in recent years, the theory of optimum reception methods began to be used widely, not only in radio communications, but also in radar, radio control, and radio astronomy, i. e., in fields where reliable signal detection is just as vital as is precise reproduction of messages these signals carry.

Initially, we will examine simple binary detection, i. e., detection of the only possible signal. Then, we will look into complex binary detection, i. e, detection of one of many possible signals (there is no requirement here to determine exactly which of m possible signals is present. It suffices to establish that one possible signal actually is present).

Simple binary detection in radar corresponds to detection of an object (aircraft) in one predetermined elementary cell of space. If the requirement is to establish whether an object is present in any one of m elementary cells (without specifying exactly which one), then this task equates to complex binary detection (in future, for brevity, the word "simple" often is omitted from the term "simple binary detection," but the word "complex" always is retained when we refer to complex binary detection.

Simultaneous detection and recognition (distinction) of m possible signals are examined in Chapters 11 and 12.

Analog message reception is examined in Chapter 13. However, here the terminology used differs somewhat from the Kotel'nikov terminology used in Part II of this book.

Kotel'nikov described the parameter of the signal carrying usable message x in the form

$$\eta = \eta_0 (1 + mx), \tag{8.1}$$

where x -- normalized message, i. e., dimensionless magnitude in the range of ± 1. For example, for AM

$$a = a_0 (1 + mx),$$

where a -- signal amplitude, while m -- percent modulation.

An (8.1) type entry is convenient and natural in the case of radio /117
communications, when message x is transmitted by means of the corresponding modulation of the signal parameter. In the case of radar, radio astronomy, radiometry, and in several other cases, signal modulation type by the message may not be selected according to our wishes and, during reception, there usually is a requirement to measure the magnitude of this or that parameter of a signal carrying a usable message. Thus, for example, there is a requirement to measure signal amplitude a or moment of arrival τ (computed from some base time) or its frequency f. In

these cases, measured parameter η, rather than one of its elements x, included in expression (8.1) may be considered a usable message.

If modulation parameters η_0 and m are known* (or may be considered known), then measurement of parameter η is fully equivalent to measurement of normalized message x. Here, transition from measurement of parameter η to measurement of normalized message x and vice versa will be accomplished easily by using relationship (8.1).

Thus, for example, it follows from this relationship that, if $P_y(x) = f_1(x)$ and $P_y(\eta) = f_2(\eta)$ -- inverse probability densities of magnitudes x and η, then

$$\left.\begin{aligned} &P_y(x) = \frac{d\eta}{dx} f_2[\eta_0(1+mx)] = m\eta_0 f_2[\eta_0(1+mx)] \\ \text{and}\quad &P_y(\eta) = \frac{dx}{d\eta} f_1\left(\frac{\eta}{m\eta_0} - 1\right) = \frac{1}{m\eta_0} f_1\left(\frac{\eta}{m\eta_0} - 1\right). \end{aligned}\right\} \tag{8.2}$$

According to these formulas, knowing distribution $P_y(x)$, it is simple to find distribution $P_y(\eta)$ and vice versa. Further exposition mainly will apply to a non-normalized message; however, in future it often is designated x, rather than η.

In several cases, there may be a requirement for simultaneous measurement of several signal parameters. For example, if it is necessary to measure both range to an airborne aircraft and its radial velocity, then the task boils down to simultaneous measurement of the time a signal reflected from an aircraft arrives and signal frequency. An analogous task arises also in the case of multichannel communications, i. e., during simultaneous transmission of several different messages.

---

*It follows from (8.1) that $\eta_0 = \frac{(\eta_{макс} + \eta_{мин})}{2}$ and $m = \frac{(\eta_{макс} - \eta_{мин})}{2}$; therefore, if limits $\eta_{мин}$ and $\eta_{макс}$ of parameter η change are known (and finite), then magnitudes η_0 and m also are known.

Measurement of just one parameter is examined in Part III for the purpose of simplicity. The special features of simultaneous measurement of two or more parameters are analyzed in Part IV of the book.

8.2 Methodolgy of Computation of Inverse Probabilities /118

Finding distribution $P_y(x)$ for precisely-known signals was examined in the preceding part of the book. We will explain which special features in $P_y(x)$ computation arise when the signal has parasitic random parameters α_1, α_2, . . ., α_n.

Initially, for simplicity, let the signal have only one parasitic random parameter α . Here, in accordance with (1.9)

$$y(t) = u_{x,\alpha}(t) + u_m(t), \tag{8.3}$$

where $u_m(t)$ -- additive noise with known a priori distribution $W_m(u_m)$.

In an analogy with relationship (4.6), in this case it is possible to write

$$P_y(x, \alpha) = kP(x, \alpha) P_{x,\alpha}(y), \tag{8.4}$$

since now the signal has two random parameters (x and α), rather than one. In this expression, $P(x, \alpha)$ -- a priori joint distribution x and α , k -- constant factor (not dependent on x and α), which may be determined from normality condition

$$\int_{A_x}\int_{A_\alpha} P_y(x, \alpha)\, dx\, d\alpha = 1, \tag{8.5}$$

where A_x and A_α -- fields of all possible values of parameters x and α , respectively, $P_y(x, \alpha)$ -- probability (probability density) that, for given realization y(t), signal random parameters equal x and α , respectively.

We are interested in probability $P_y(x)$, i. e., the probability that, for given realization y(t), the message equals x; here, parasitic parameter α may

have any value. Therefore, $P_y(x)$ determination requires integration of function $P_y(x, \alpha)$ with respect to all possible parameter α values:

$$P_y(x) = \int_{A_\alpha} P_y(x, \alpha)\, d\alpha. \tag{8.6}$$

It follows from formulas (8.4) and (8.6) that finding distribution $P_y(x)$ requires knowing functions $P_y(x, \alpha)$ and $P_{x\alpha}(y)$.

In the general case

$$P(x, \alpha) = P(x) P_x(\alpha); \tag{8.7}$$

if x and α statistically independent, which usually is the case, then

$$P(x, \alpha) = P(x) P(\alpha), \tag{8.8}$$

where P(x) and $P(\alpha)$ — a priori distributions of message x and parasitic /119 parameter α, respectively, which are assumed known or which must be given due to certain circumstances (see Chapter 19). Consequently, distribution $P_{x\alpha}(y)$ remains to be found.

If additive noise $u_m(t)$ has no statistical coupling with parameters x and α (which usually is the case), then it follows from (8.3) that probability $P_{x\alpha}(y)$ that total oscillation equals y for given x and α is determined from the following relationship:

$$P_{x,\alpha}(y) = W_m[y(t) - u_{x,\alpha}(t)], \tag{8.9}$$

i. e., distribution $P_{x\alpha}(y)$ is obtained from noise distribution $W_m(u_m)$ by replacement of argument u_m by difference $(y - u_{x\alpha})$. Thus, for example, for normal white noise, in accordance with formulas (1.25) and (8.9), we obtain

$$P_{x,\alpha}(y) = \frac{1}{(\sqrt{2\pi N})^n} \exp\left[-\frac{1}{N_0}\int_0^T [y(t) - u_{x,\alpha}(t)]^2\, dt\right]. \tag{8.10}$$

Thus, in a case of signals with random parameters, distribution $P_y(x)$ computation requires that:

1. Distribution $P_{x\alpha}(y)$ with respect to formula (8.9) be computed with respect to known additive noise distribution $W_m(u_m)$.

2. Distribution $P_y(x, \alpha)$ be computed with respect to formulas (8.4) and (8.7) [or (8.8)].

3. Desired distribution $P_y(x)$ be computed with respect to formula (8.6).

After distribution $P_y(x)$ is found, further mathematical analysis is the same as analysis of a precisely-known signal.

For example, if the optimum receiver operates on the maximum inverse probability principle, then oscillation $\dot{y}(t)$ at receiver input is determined by relationship (4.2).

If the signal has n parasitic parameters $\alpha_1, \alpha_2, \ldots, \alpha_n$, rather than one parameter, then, analogous to the relationships presented above, the result is

$$P_y(x)=\int_{A_{\alpha_1}} \ldots \int_{A_{\alpha_n}} P_y(x;\ \alpha_1,\ \ldots,\ \alpha_n)\, d\alpha_1 \ldots d\alpha_n, \tag{8.11}$$

where

$$P_y(x;\ \alpha_1,\ \ldots,\ \alpha_n)=kP(x;\ \alpha_1,\ \ldots,\ \alpha_n)\, P_{x;\ \alpha_1,\ \ldots,\ \alpha_n}(y); \tag{8.12}$$

$$P_{x;\ \alpha_1,\ \ldots,\ \alpha_n}(y)=W_m[y(t)-u(x;\ \alpha_1,\ \ldots,\ \alpha_n;\ t)]. \tag{8.13}$$

Given independent x, $\alpha_1, \ldots, \alpha_n$

$$P(x;\ \alpha_1,\ \ldots,\ \alpha_n)=P(x)\, P(\alpha_1) \ldots P(\alpha_n). \tag{8.14}$$

CHAPTER NINE /120

BINARY DETECTION

9.1 Computation of the Inverse Probability of a Random-Phase Signal

Let

$$y(t)=u_{a,\varphi}(t)+u_{\text{ш}}(t), \tag{9.1}$$

where

$$u_{a,\varphi}(t)=a\cos(\omega t+\varphi). \tag{9.2}$$

We will compute distribution $P_y(a)$ of signal amplitude inverse probability densities, assuming frequency ω is known, while phase φ is random and equally probable, i. e.

$$\left.\begin{array}{l} P(\varphi)=\frac{1}{2\pi}, \text{ where } \quad \varphi=0\div 2\pi; \\ P(\varphi)=0 \text{ outside these limits} \end{array}\right\} \tag{9.3}$$

Amplitude a has known a priori distribution P(a). Amplitude a and phase ϕ are assumed to be analog random magnitudes, i. e., during observation cycle (0, T), they are constant, but random, and have probability densities P(a) and

$P(\phi)$. Distribution $P_y(a)$ computed in this section in future will be used both for solution of the binary detection problem and for signal amplitude measurement.

During distribution $P_y(a)$ computation, amplitude a plays the role of desired usable signal x, while phase ϕ is a parasitic random parameter. Therefore, general formulas (8.11)--(8.14) in this case have the form

$$P_y(a) = \int_0^{2\pi} P_y(a, \varphi)\, d\varphi, \tag{9.4}$$

where

$$P_y(a, \varphi) = kP(a, \varphi)\, P_{a,\varphi}(y); \tag{9.5}$$

$$P_{a,\varphi}(y) = \frac{1}{(\sqrt{2\pi N})^n} \exp\left[-\frac{1}{N_0}\int_0^T [y(t) - u_{a,\varphi}(t)]^2\, dt\right]; \tag{9.6}$$

$$P(a, \varphi) = P(a)\, P(\varphi) \tag{9.7}$$

(a and ϕ are assumed to be statistically-independent magnitudes).

Expanding the expression (9.6) parentheses under the integral and considering relationship (9.2), we obtain /121

$$P_{a,\varphi}(y) = \frac{1}{(\sqrt{2\pi N})^n} e^{-\frac{1}{N_0}\int_0^T y^2(t)\, dt} e^{-\frac{a^2 T}{2N_0}} e^{\eta(a,\varphi)}, \tag{9.8}$$

where

$$\eta(a, \varphi) = \frac{2a}{N_0}\int_0^T y(t)\cos(\omega t + \varphi)\, dt. \tag{9.9}$$

From relationships (9.3)--(9.5), (9.7), and (9.8), we obtain

$$P_y(a) = k_1 P(a)\, e^{-\frac{a^2 T}{2N_0}} \int_0^{2\pi} e^{\eta(a,\varphi)}\, d\varphi, \tag{9.10}$$

where k_1 — constant, which includes all factors not dependent on variable a.

It follows from (9.9) that

$$\eta(a, \varphi) = \frac{2a}{N_0}(X\cos\varphi - Y\sin\varphi), \tag{9.11}$$

where

$$X = \int_0^T y(t)\cos\omega t\,dt; \quad Y = \int_0^T y(t)\sin\omega t\,dt. \tag{9.12}$$

To simplify expression (9.10) integration, we will introduce new variables M and θ so that

$$X = M\cos\theta; \quad Y = M\sin\theta; \text{ i. e. } M = \sqrt{X^2+Y^2}. \tag{9.13}$$

Then, expression (9.11) takes the form

$$\eta(a, \varphi) = \frac{2a}{N_0}M\cos(\theta+\varphi), \tag{9.14}$$

and the integral included in formula (9.10) easily will be removed:

$$\begin{aligned} P_y(a) &= k_1 P(a)\,e^{-\frac{a^2T}{2N_0}} \int_0^{2\pi} e^{\frac{2aM}{N_0}\cos(\theta+\varphi)}\,d\varphi = \\ &= k_2 P(a)\,e^{-\frac{a^2T}{2N_0}} I_0\left(\frac{2aM}{N_0}\right), \end{aligned} \tag{9.15}$$

where $I_0(z)$ -- modified Bessel function.

Since it is important to know only the curve $P_y(y)$ shape, and its scale plays no principal role, then it is possible to assume that $k_2 = 1$ in (9.15). In addition, it often is more advisible during computational device realization to compute /122 ln $P_y(a)$ rather than $P_y(a)$. Therefore, instead of (9.15), it also is possible to write

$$\ln P_y(a) = \ln P(a) - \frac{a^2T}{2N_0} + \ln I_0\left(\frac{2aM}{N_0}\right). \qquad (9.15a)$$

It follows from expression (9.15) that magnitude M must be computed prior to determination of desired distribution $P_y(a)$. In the general case, this parameter may be computed using formulas (9.12) and (9.13).

A functional diagram of the corresponding computational device is depicted in Figure 9.1. It will comprise generators $Г_1$ and $Г_2$, which generate the sinu-

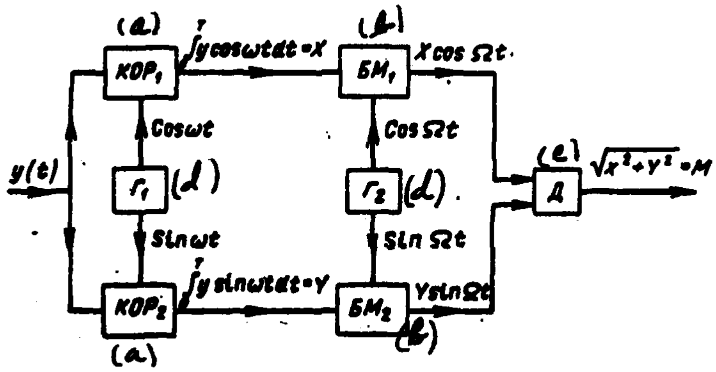


Figure 9.1. (a) -- KOR [correlator]; (b) -- BM [balanced modulator]; (c) -- D [detector]; (d) -- G [generator].

soidal voltages of signal ω frequency and some random supplemental frequency Ω, correlators KOP_1 and KOP_2, balance modulators $БM_1$ and $БM_2$, which perform multiplication operations, and linear amplitude detector $Д$, which separates the total oscillation $X\cos\Omega t + Y\sin\Omega t$ envelope. The magnitude of this envelope also equals desired parameter M (all constant factors in Figure 9.1 for simplicity are assumed to equal unity).

A computational device corresponding to the Figure 9.1 circuit is relatively complex. However, given several assumptions which usually are the case, it may

be replaced by an optimum linear filter, which provides maximum signal-to-noise ratio and which is described in § 2.3.

Actually, it follows from formulas (9.9) and (9.14) that M is the envelope of oscillation $\frac{N_0}{2a}\eta(a, \varphi)$, considered a function of φ, i. e., the envelope of function

$$\beta(\varphi) = \int_0^T y(t)\cos(\omega t + \varphi)\,dt. \tag{9.16}$$

But, it was shown in § 4.3 that, if signal $u_c(t)$ lasts only during interval (0, T), then optimum filter output voltage at moment T equals /123

$$u_{\text{вых}}(T) = \text{const}\int_0^T y(t)\,u_c(t)\,dt. \tag{9.17}$$

Here, optimum is understood to mean a filter matched with signal $u_c(t)$, i. e., having impulse transient characteristic equalling

$$\eta(t) = \text{const}\,u_c(t_0 - t),$$

where $t_0 = T$.

Comparing expressions (9.16) and (9.17), it is easy to become convinced that $\beta(\varphi)$ coincides (precise to a constant factor) with the voltage magnitude (at moment T) at the output of an optimum linear filter matched with signal $u_{a\,\phi}(t)$.*

In accordance with (9.16), this voltage is a function of rf occupation phase

*We considered that the filter must be matched with a signal in the form $\text{const}\cdot\cos(\omega t + \varphi)$. Phase ϕ of this signal is random and, consequently, is unknown. But, since the requirement is reproduction of just the envelope of the output voltage of this filter, then any phase ϕ, such as $\phi = 0$, may be selected during filter design.

Φ . But, magnitude M is the function $\beta(\tau)$envelope. Therefore, it is proportional to the value of the envelope (at moment T) of the optimum linear filter output voltage.

It is evident that confirmation is sufficiently precise if the envelope exists at filter output, i. e. if $\omega T \gg 1$, and, during observation cycle T, at least several rf occupation periods accumulate. This condition usually is met.

Consequently, as the optimum receiver is realized, magnitude M included in expression (9.15) usually may be obtained by oscillation y(t) passage through

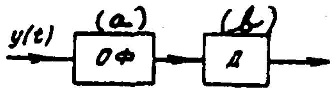


Figure 9.2. (a) -- OF [optimum filter];
(b) -- D [detector].

matched optimum linear filter ОФ and separation of the oscillation envelope at filter output (at moment T) with the aid of amplitude detector Д (Figure 9.2).

Realization of the remaining operations required to compute function ln $P_y(a)$ in accordance with (9.15a) does not encounter major difficulties. In particular, the most complicated of these operations, computation of an ln $I_0(z)$ type function, may be accomplished by using a conventional thermionic diode detector. Actually, it was shown in [112] that a conventional diode detector (thermionic diode) for large load impedances (R $\geqslant$10 - 50 kilohms) has a response curve of the type

$$\Delta U_{=} = \frac{1}{b} \ln I_0 (bU_m). \qquad (9.18)$$

where U_m -- sinusoidal oscillation amplitude at detector output; $\Delta U_{=}$ -- rise /124 in the direct detector output voltage component caused by onset of an oscillation with amplitude U_m;

$$b = \mathrm{const} \approx 10\ 1/\mathrm{B}.$$

The greater detector load impedance R, the more precisely its response curve is determined by expression (9.18).

Thus, realization of the optimum receiver computing inverse probability $P_y(a)$ or its logarithm does not encounter major difficulties.

9.2 Binary Detection of a Random Initial Phase Signal

Let the signal at receiver output either be absent or have the form

$$u_c(t) = a_0 \cos(\omega t + \varphi), \tag{9.19}$$

where amplitude a_0 and frequency ω are precisely known, but phase ϕ is random and equally probable, i. e., distributed with respect to law (9.3). Evidently, solution of the problem of detecting this signal equates to the response to the question of which of two possible values of amplitude a occur during a given observation cycle: $a = a_0$ or $a = 0$.

In accordance with the inverse probability approach, answering this question requires comparison between themselves of inverse probabilities $P_y(a_0)$ and $P_y(0)$ of the corresponding amplitude values.

It follows from general expression (9.15) that

$$\left.\begin{aligned} P_y(a_0) &= k_2 P(a_0)\, e^{-a_0^2 T/2N_0}\, I_0\left(\frac{2a_0 M}{N_0}\right); \\ P_y(0) &= k_2 P(0). \end{aligned}\right\} \tag{9.20}$$

If the maximum inverse probability principle is used, then the value of the amplitude with the greatest inverse probability should be considered the decision. In this case, this denotes that a signal present decision ($a_0 = a$) must be used if it turns out that

$$P_y(a_0) > P_y(0), \qquad (9.21)$$

and the signal absent decision (a = 0) in the opposite case, i. e.

$$P_y(a_0) \leqslant P_y(0). \qquad (9.22)$$

It follows from (9.20) and (9.21) that a "yes" decision ("signal"), i. e., (a_0 = 0) must be used when satisfying inequality

$$\ln I_0\left(\frac{2a_0 M}{N_0}\right) > U_0, \qquad (9.23)$$

where /125

$$U_0 = \frac{a_0^2 T}{2N_0} + \ln\frac{P(0)}{P(a_0)}. \qquad (9.24)$$

Consequently, the optimum receiver must compute magnitude $\ln I_0(2a_0M/N_0)$ and compare it with some threshold U_0. The response "yes" (signal) must be supplied if the threshold is exceeded, with the response "no" (noise only, no signal) provided in the opposite case.

Here, determination of parameter M included in expression (9.23) may be accomplished as indicated in the preceding section, i. e., through use of the Figure

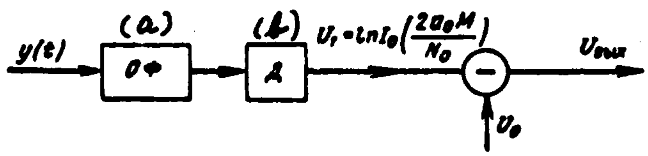


Figure 9.3. (a) -- OF [optimum filter]; (b) -- D [detector].

9.1 computational device or with the aid of an optimum linear filter (Figure 9.2). In the latter case, the optimum detector circuit has the form shown in Figure 9.3.

If $U_1 > U_0$ (i. e., $U_{вых} > 0$) is obtained at detector $Д$ output at moment $t = T$, the response is "yes," while it is "no" in the opposite case.*

Instead of a detector with an $\ln I_0(z)$ curve, one with any other monotonic response curve may be used in the Figure 9.3 circuit, if threshold U_0 magnitude is adjusted accordingly.

Actually, if voltages U_1 and U_0 are subjected to monotonic nonlinear transformation with respect to law $U' = f(U)$ (Figure 9.4), then, given any (but monotonic)

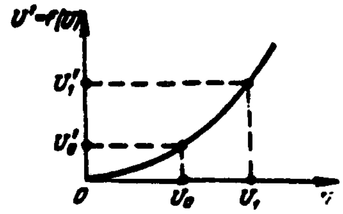


Figure 9.4

type of this transformation, inequalities $U_1 > U_0$ and $U_1 < U_0$ will be transformed into inequalities $U' > U_0'$ and $U_1' < U_0'$, respectively. Therefore, such a transformation does not impact upon solution of the signal/no signal problem. Consequently, when response curve shape changes, accompanied by the corresponding threshold U_0 change, signal detection error probablities will remain unchanged.

As pointed out in § 5.2, two types of errors occur in binary detection--false alarms, with probability $P_{лт}$, and signal misses, with probability /126 $P_{п}$. Here, composite error probability equals

$$P_{ош} = P(0)\,P_{лт} + P(a_0)\,P_{пр}. \tag{9.25}$$

The methodology for computation of error probabilities $P_{лт}$ and $P_{пр}$ in this case is analogous to that presented in § 5.2.

*In Figure 9.3 and subsequently, the device clipping detector output voltage at moment $t = T$ is assumed to be included in unit $Д$.

We will examine, for example, computation of probability $P_{\text{n T}}$.

It follows from (9.23) that false-alarm probability equals the probability of inequality (9.23) satisfaction for an absent signal, i. e., when only noise is present.

Inequality (9.23) may be written in the form

$$\frac{2a_0 M}{N_0} > z, \tag{9.26}$$

where z is determined from the relationship

$$\ln I_0(z) = U_0. \tag{9.27}$$

It follows from (9.13) and (9.26) that a false alarm will occur when the following inequality is satisfied:

$$\eta^2 > z^2, \tag{9.28}$$

where

$$\eta^2 = \left(\frac{2a_0}{N_0} X\right)^2 + \left(\frac{2a_0}{N_0} Y\right)^2. \tag{9.29}$$

It follows from (9.12) that, for a signal, i. e., when y(t) = $u_m(t)$, these relationships occur

$$\left.\begin{aligned} \frac{2a_0}{N_0} X &= \frac{2}{N_0}\int_0^T u_m(t)\, a_0 \cos\omega t\, dt; \\ \frac{2a_0}{N_0} Y &= \frac{2}{N_0}\int_0^T u_m(t)\, a_0 \sin\omega t\, dt. \end{aligned}\right| \tag{9.30}$$

It was demonstrated in § 5.2 that random magnitude $\frac{2}{N_0}\int_0^T u_m(t)\, u_c(t)\, dt$

has a normal law of distribution with a zero mean value and dispersion $2Q/N_0$, where $Q=\int_0^T u_c^2(t)\,dt$.

Therefore, random magnitudes $\frac{2a_0}{N_0}X$ and $\frac{2a_0}{N_0}Y$, determined by expressions (9.30), also have normal laws of distribution with zero mean values and identical dispersions equalling

$$\sigma^2=\frac{2Q_0}{N_0}, \tag{9.31}$$

where /127

$$Q_0=\frac{a_0^2 T}{2}. \tag{9.32}$$

In addition, considering orthogonality of functions cos ωt and sin ω t included in expressions (9.30), it is easy to become convinced that magnitudes $\frac{2a_0}{N_0}X$ and $\frac{2a_0}{N_0}Y$ are statistically independent. Consequently, magnitude η^2 is the sum of the squares of two normally-distributed independent random magnitudes with zero mean values and identical dispersions σ^2 . Such a random magnitude is subordinate, as is known, to distribution χ^2 with two degrees of freedom, i. e.,

$$P(\eta^2>z^2)=e^{-z^2/2\sigma^2}. \tag{9.33}$$

But, it follows from (9.28) that $P(\eta^2 > z^2)$ is nothing but a false-alarm probability. Therefore

$$P_{\text{лт}}=e^{-z^2/2\sigma^2}$$

or, considering (9.31)

$$P_{\text{лт}}=e^{-(N_0/4Q_0)z^2}, \tag{9.34}$$

where, in accordance with formulas (9.24), (9.27), and (9.32), magnitude z is determined by the relationship

$$\ln I_0(z) = U_0 = \frac{Q_0}{N_0} + \ln\frac{P(0)}{P(a_0)}. \tag{9.35}$$

It follows from formulas (9.34) and (9.35) that it suffices for computation of a false-alarm probability to know signal-to-noise ratio Q_0/N_0 and ratio $P(0)/P(a_0)$ of a priori signal/no signal probabilities (or threshold U_0).

Miss probability P_{np} is determined by expression

$$P_{np} = 1 - \frac{N_0}{2Q_0} e^{-Q_0/N_0} \int_z^\infty x e^{-(N_0/4Q_0)x^2} I_0(x)\,dx, \tag{9.36}$$

which may be obtained using various approaches. We will examine one such approach now. A second approach will be presented in Part IV (§ 14.6).

Beforehand, we replace z with α in formulas (9.33) and (9.36), where

$$z = \sqrt{\frac{2Q_0}{N_0}}\,\alpha, \tag{9.37}$$

then, we will obtain:

$$P_{лт} = e^{-\alpha^2/2}; \tag{9.38}$$

$$P_{np} = 1 - e^{-Q_0/N_0} \int_\alpha^\infty y e^{-y^2/2} I_0\left(\sqrt{\frac{2Q_0}{N_0}}\,y\right) dy. \tag{9.39}$$

In derivation of formula (9.39), we use the aforementioned assumption /128 whereby error probabilities $P_{л}$ and P_{np} will not change if the shape of response curve $\ln I_0\left(\frac{2a_0M}{N_0}\right)$ in the Figure 9.3 circuit is replaced by any other monotonic function M and the threshold value simultaneously is adjusted from U_0 to some other magnitude U_0'. Therefore, instead of Figure 9.3, we may examine the Figure

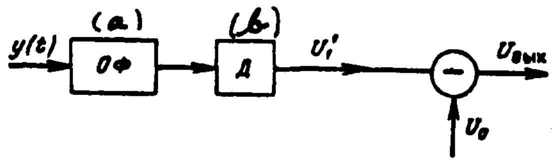


Figure 9.5. (a) -- OF [optimum filter]; (b) -- D [detector].

9.5 circuit. Here, Д -- linear amplitude detector with a single transfer constant, which at moment T clips the optimum filter ОФ output voltage envelope $U_m(t)$ value, i. e.

$$U_1' = U_m(T). \tag{9.40}$$

If U_1' U_0' results, the response is "yes" (signal); otherwise the response is "no" (no signal).

Consequently, false-alarm probability $P_{лт}$ is the probability that, at moment T, noise envelope $U_{mш}(T)$ at linear filter output will exceed U_0'.

Since noise at filter output has normal distribution, then, in view of filter linearity, the noise retains normal distribution at filter output also and the envelope of this noise $U_{mш}(t)$ is subordinate to Rayleigh's law. Therefore,

$$P_{лт} = P[U_{mш}(T) > U_0'] = e^{-U_0'^2/2U_ш^2}, \tag{9.41}$$

where $U_ш^2$ -- mean square of noise voltage $u_ш(t)$ at filter output.

Miss probability equals

$$P_{пр} = 1 - P_{по}, \tag{9.42}$$

where $P_{по}$ -- probability of correct signal detection given the condition that

there is a signal. Consequently, $P_{по}$ is the probability that the signal-plus-noise envelope at moment T will exceed threshold U_0'

$$P_{по} = P[U_{mcш}(T) > U_0'], \qquad (9.43)$$

where $U_{mcш}$ — value of the envelope (at moment T) of an oscillation, which is the sum of signal sinusoidal voltage and normal noise voltage.

It is known [21] that the probability density of the envelopes of such an oscillation is determined by expression

$$P(U_{mcш}) = \frac{U_{mcш}}{U_ш^2} \exp\left(-\frac{U_{mcш}^2 + U_{mc}^2}{2U_ш^2}\right) I_0\left(\frac{U_{mc} U_{mcш}}{U_ш^2}\right),$$

where U_{mc} — signal amplitude value. Therefore,

$$P_{по} = \int_{U_0'}^{\infty} P(U_{mcш})\, dU_{mcш} = \int_{U_0'}^{\infty} \frac{U_{mcш}}{U_ш^2} \times$$

$$\times \exp\left(-\frac{U_{mc}^2 + U_{mcш}^2}{2U_ш^2}\right) I_0\left(\frac{U_{mc} U_{mcш}}{U_ш^2}\right) dU_{mcш}.$$

Designating $U_{mcш}/U_ш = y$, we obtain

$$P_{по} = e^{-U_{mc}^2/2U_ш^2} \int_{U_0'/U_ш}^{\infty} y e^{-y^2/2} I_0\left(\frac{U_{mc}}{U_ш} y\right) dy. \qquad (9.44)$$

Here, $U_{mc}/U_ш$ is the ratio of signal amplitude to the mean square value of the noise voltage at optimum filter output at moment T, i. e., at that moment when signal voltage is maximum. It was shown in § 2.3 that this ratio equals $\sqrt{2Q/N_0}$ [see formula (2.34)], i. e., in this case

$$\frac{U_{mc}}{U_ш} = \sqrt{\frac{2Q_0}{N_0}}. \qquad (9.45)$$

Therefore, formula (9.44) may be written in the form

$$P_{\text{по}} = e^{-Q_0/N_0} \int\limits_{U_0'/U_{\text{ш}}}^{\infty} y e^{-y^2/2} I_0\left(\sqrt{\frac{2Q_0}{N_0}} y\right) dy. \tag{9.46}$$

Considering (9.42), we have

$$P_{\text{пр}} = 1 - e^{-Q_0/N_0} \int\limits_{U_0'/U_{\text{ш}}}^{\infty} y e^{-y^2/2} I_0\left(\sqrt{\frac{2Q_0}{N_0}} y\right) dy. \tag{9.47}$$

It follows from comparison of expressions (9.38) and (9.41) that

$$\frac{U_0'}{U_{\text{ш}}} = \alpha; \tag{9.48}$$

therefore, expression (9.47) coincides with (9.39), which also required proof.

Thus, it is demonstrated that formulas (9.38) and (9.39) determine false-alarm and miss probabilities during detection of a sinusoidal signal with an equally-probable phase (on a normal white noise background). Curves providing the link

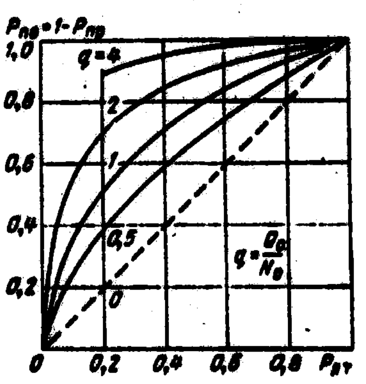


Figure 9.6

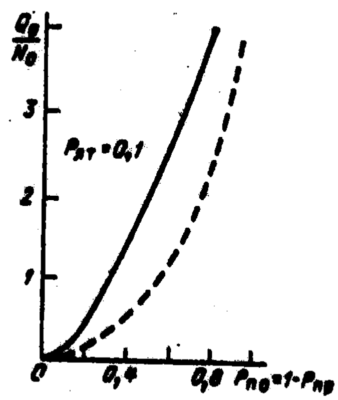


Figure 9.7

between error probabilities $P_{л}$, and P_{np} for various signal-to-noise Q_0/N_0 values

may be computed using these formulas (Figure 9.6). These curves were computed for the first time in work done by Peterson and others [17].

There is no requirement to use the curves for slight error probabilities since simple analytical expressions may be obtained for these cases.

It was demonstrated in [106] that, given $P_{лт} \leqslant 0{,}1$ and $P_{пр} \leqslant 0{,}1$, from formulas (9.38) and (9.39) one obtains the following with less than a 0.5 дБ loss

$$\frac{Q_0}{N_0} = \left(\sqrt{\ln \frac{1}{P_{лт}}} + \sqrt{\ln \frac{1}{P_{пр}} - 1{,}4}\right)^2. \tag{9.49}$$

Where $P_{пр} \to 0$ and $P_{лт} \to 0$, the error for this formula asymptotically will strive towards zero.

It follows from comparing formula (9.49) with corresponding formula (5.31a), obtained for a precisely-known signal, that, given $P_{лт} \leqslant 0{,}1$ and $P_{пр} \leqslant 0{,}1$, they essentially coincide.

Consequently, given $P_{лт} \leqslant 0{,}1$ and $P_{пр} \leqslant 0{,}1$, approximately the same signal energy is required to detect a random-phase signal as is the case for a precisely-known signal.

This conclusion already is invalid for high error probabilities. Thus, for example, relationships of Q_0/N_0 to Q_0/N_0 where $P_{лт}$ = 0.1 are depicted in Figure 9.7 for a random-phase (solid curve) and a precisely-known (dotted curve) signal. It is evident from comparing these curves that, where $P_{лт}$ = 0.1 and $P_{пр}$ = 0.5, the energy required to detect a random-phase signal is greater by a factor of almost 2.5 than is the case for a precisely-known signal and, given a further /131 decrease in $P_{по}$ magnitude, this difference will become even more pronounced.

We now will examine in more detail the problem of threshold bias selection at detector output. Detector output and threshold bias in the Figure 9.3 circuit are characterized by dimensionless magnitudes $U_1 = \ln I_0\left(\frac{2a_0 M}{N_0}\right)$ and U_0, where expression (9.24) determines U_0.

In the corresponding real circuit, there is a requirement to separate the matched linear filter output voltage envelope and compare it at moment t = T with some real threshold U_{op} (having voltage dimensionality). Therefore, if one assumes, as was the case when examining the Figure 9.5 circuit, that a linear detector with a single transfer constant separates the envelope, then

$$U_{op} = U_0',$$

and, in accordance with formulas (9.38) and (9.48), we obtain

$$U_{op} \approx U_{ш} \sqrt{2 \ln \frac{1}{P_{лт}}},$$

where $U_{ш}$ -- extant envelope noise voltage value at detector output, i. e., at optimum linear filter output.

Since

$$\frac{u_{c\,max}(T)}{U_{ш}} = \sqrt{\frac{2Q}{N_0}} = \sqrt{2q},$$

one also may assume that

$$U_{op} \approx u_{c\,max}(T) \frac{1}{\sqrt{2q}} \sqrt{2 \ln \frac{1}{P_{лт}}}.$$

The aforementioned analysis was performed for a signal changing with respect to law (9.2), i. e., for an unmodulated sinusoidal signal. If the signal has a more common form, one modulated with respect to amplitude and with respect to phase (frequency), i. e.,

$$u_c(t) = a(t) \cos[\omega t + \psi(t) + \varphi], \tag{9.50}$$

where a(t) and ψ(t) -- precisely-known time functions, then it is possible to show [17] that all aforementioned results will remain valid, with the following provisos:

a) signal energy Q_0 should be replaced by Q, where

$$Q=\frac{1}{2}\int_0^T a^2(t)\,dt;$$

b) the inverse probablity of signal presence equals $P_y(C)$, rather than $P_y(a_0)$ [formula (9.20)], where /132

$$\left.\begin{aligned}
&P_y(C)=k_2P(C)\,e^{-Q/N_0}I_0\left(\frac{2}{N_0}M_1\right);\\
&M_1=\sqrt{X_1^2+Y_1^2};\\
&X_1=\int_0^T y(t)\,a(t)\cos[\omega t+\psi(t)]\,dt;\\
&Y_1=\int_0^T y(t)\,a(t)\sin[\omega t+\psi(t)]\,dt,
\end{aligned}\right\}\tag{9.51}$$

where P(C) -- a priori probability of signal presence.

It follows from these relationships that parameter M_1, just as was the case for the parameter M mentioned previously, may be computed with the aid of an optimum filter and amplitude detector or by the Figure 9.1 computing device. In the latter case, cos ωt and sin ωt should be replaced everywhere they appear by a(t) cos [ωt + $\psi(t)$] and a(t) sin [ωt + $\psi(t)$], respectively.

Optimum filter structure also is complicated accordingly.

9.3 Binary Detection of a Fluctuating Signal

Let the signal have the form

$$u_c(t)=a\cos(\omega t+\varphi),\tag{9.52}$$

where frequency ω is precisely known, while amplitude a and initial phase φ during observation cycle T are unchanged, but fluctuate from one observation (observation cycle) to another, i. e., they change as analog random magnitudes with some distributions W(a) and P (φ), respectively. For brevity, this is called a fluctuating signal.

We will assume that distribution P(ϕ) is uniform, i. e., it is described by relationships (9.3).

Let the a priori probabilities of the presence and absence of such a signal equal P(C) and P(0), respectively. If there is a signal at input, then its amplitude has some non-zero value a. If there is no signal at input, then this denotes that its amplitude is identical with zero ($a \equiv 0$). Therefore, solution of the signal detection problem means answering the question of whether or not amplitude a has some non-zero value (it is unimportant exactly what the value is) or if it is identical with zero.

In accordance with the inverse probability approach, the answer to this question requires comparison between themselves of inverse probabilities $P_y(C)$ and $P_y(0)$, where $P_y(C)$ -- inverse probability that amplitude a has some non-zero value (it is unimportant exactly what value), while $P_y(0)$ -- inverse probability that amplitude a is identical with zero.

It follows from formula (9.15) that, for a random-phase equiprobable /133 signal, amplitude a inverse probability density $P_y(a)$ equals

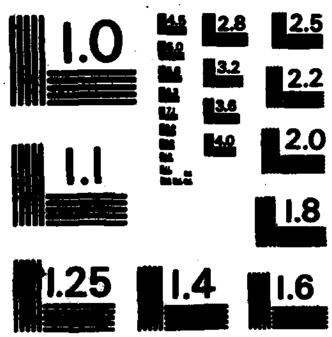
$$P_y(a) = k_1 P(a) e^{-a^2 T/2N_0} I_0\left(\frac{2aM}{N_0}\right). \qquad (9.15)$$

Since $P_y(C)$ is the inverse probability that amplitude a has any non-zero value, then

$$P_y(C) = \int_0^\infty P_y(a)\,da, \qquad (9.53)$$

AD-A120 899 THEORY OF OPTIMUM RADIO RECEPTION METHODS IN RANDOM NOISE(U) FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OH L S GUTKIN 24 SEP 82 FTD-ID(RS)T-0784-82

UNCLASSIFIED F/G 9/4 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

i. e.

$$P_y(C) = k_2 \int_0^\infty P(a) e^{-a^2T/2N_0} I_0\left(\frac{2aM}{N_0}\right) da, \tag{9.54}$$

where P(a) -- a priori non-zero amplitude a probability density. Therefore, in this case

$$P(a) = P(C) W(a) \tag{9.55}$$

and

$$P_y(C) = k_2 P(C) \int_0^\infty W(a) e^{-a^2T/2N_0} I_0\left(\frac{2aM}{N_0}\right) da. \tag{9.56}$$

It follows from (9.15) that

$$P_y(0) = k_2 P(0). \tag{9.57}$$

Coefficient k_2 included in formulas (9.56) and (9.57) is determined from a normality condition which, for the case being examined, has the form

$$P_y(C) + P_y(0) = 1. \tag{9.58}$$

For simplicity, in future we will proceed from the maximum inverse probability criterion, i. e., assume that there is a signal, if

$$P_y(C) > P_y(0), \tag{9.59}$$

and there is no signal if

$$P_y(C) \leqslant P_y(0). \tag{9.60}$$

In addition, for specificity we will assume that signal amplitude distribution W(a) is subordinate to Rayleigh's law, i. e.,

$$W(a) = \frac{a}{\overline{u_c^2}} e^{-a^2/2\overline{u_c^2}}. \qquad (9.61)$$

Here, expression (9.56) takes the form /134

$$P_y(C) = k_2 P(C) \frac{1}{\overline{u_c^2}} \int_0^\infty a e^{-a^2/2b} I_0\left(\frac{2Ma}{N_0}\right) da, \qquad (9.62)$$

where

$$\frac{1}{b} = \frac{1}{\overline{u_c^2}} + \frac{T}{N_0} = \frac{1}{\overline{u_c^2}}\left(1 + \frac{Q_{cp}}{N_0}\right); \qquad (9.63)$$

$$Q_{cp} = \overline{u_c^2}\, T \qquad (9.64)$$

is average signal specific energy.

The integral in expression (9.62) is tabular and has the following value:

$$\int_0^\infty a e^{-a^2/2b} I_0\left(\frac{2Ma}{N_0}\right) da = b e^{(2b/N_0^2) M^2}.$$

Substituting it into (9.62), we obtain

$$P_y(C) = k_2 P(C) \frac{b}{\overline{u_c^2}} e^{(2b/N_0^2) M^2}. \qquad (9.65)$$

It follows from relationships (9.57), (9.59), and (9.65) that a detector must supply the response "yes" (signal at input) when the following inequality is satisfied

$$M > u_{0,p}, \qquad (9.66)$$

where

$$u_{0\Phi} = \sqrt{\frac{N_0^2}{2b} \ln \left[\frac{P(0)}{P(C)} \cdot \frac{\overline{u_c^2}}{b} \right]}. \tag{9.67}$$

If inequality (9.66) is not satisfied, then the decision is that there is no signal. Consequently, the task of the optimum fluctuating signal detector boils down to parameter M computation and its comparison with threshold $u_{0\Phi}$.

It was shown in the preceding section that magnitude M is proportional to the value (at moment T) of the envelope of the oscillation at optimum linear filter output. Therefore, as was true for a signal of known amplitude, the optimum fluctuating signal detector has the form depicted in Figure 9.5. All this circuit's parameters also will remain unchanged, with the exception of threshold bias magnitude.

This optimum receiver coincidence (except for threshold bias) for cases of fluctuating and non-fluctuating signal amplitudes is explained by the fact that the receiver's optimum filter $O\Phi$ is a linear system whose parameters are independent of signal amplitude; the shape of the envelope detector response curve, as noted above, also plays no role, if the threshold magnitude is adjusted accordingly /135 when it changes.

Hence, it follows that the transition from optimum detection of a signal with precisely-known amplitude to optimum detection of a random-amplitude signal subordinate to any distribution law W(a) may be accomplished without any detector (receiver) circuitry changes, except for the corresponding adjustment of threshold bias U_0' at output.

For a Rayleigh law of amplitude distr. [formula (9.61)], the threshold bias magnitude, reduced to function $\beta(\varphi)$enve... equals $u_{0\Phi}$ and is determined from formula (9.67). However, in a real circuit (Figure 9.5), envelope detector output voltage is compared at moment t = T with threshold U_0'. If it is assumed that this detector's transfer constant (for the envelope) equals unity, then optimum linear filter output voltage equalling

$$u_{\text{вых}}(T) = c_{\Phi}\int_0^T y(t)\, u_c(t)\, dt = c_{\Phi}\, a \int_0^T y(t) \cos(\omega t + \varphi)\, dt. \tag{9.68}$$

is compared with threshold U_0' at moment $t = T$.

The result in the absence of noise is

$$u_{c\,\text{вых}}(T) = c_{\Phi}\int_0^T u_c(t)\, u_c(t)\, dt = c_{\Phi}\frac{a^2}{2}T,$$

therefore

$$c_{\Phi} = \frac{2}{a^2 T}\, u_{c\,\text{вых}}(T). \tag{9.69}$$

It follows from comparison of expressions (9.16) and (9.68) that voltage $u_{\text{вых}}(T)$ exceeds magnitude $\beta(\varphi)$ by a factor of $c_{\Phi}a$. Therefore, real threshold U_0' must exceed threshold $u_{0\Phi}$ determined by formula (9.67) also by factor $c_{\Phi}a$, i. e.,

$$U_{0p} = U_0' = c_{\Phi}\, a u_{0\Phi}. \tag{9.70}$$

Considering relationships (9.63), (9.64), (9.67), (9.69) and the fact that

$$\frac{u_{c\,\text{вых}}(T)}{U'_{ш}} = \sqrt{\frac{2Q}{N_0}} = a\sqrt{\frac{T}{N_0}},$$

it is possible to reduce formula (9.70) to the following form:

$$U_{0p} = \sqrt{2}\, U_{ш}\sqrt{\left(1 + \frac{1}{q}\right)\ln\left[\frac{P(0)}{P(C)}(1 + q)\right]} \tag{9.71}$$

(where $\frac{P(0)}{P(C)}(1+q) < 1$, one should assume $U_{0p} = 0$).

where

$$q = \frac{Q_{cp}}{N_0}. \tag{9.72}$$

Now, we will find fluctuating signal error probabilities $P_{\text{л т}}$, $P_{\text{пр}}$ and /136
$P_{\text{ош}}$, assuming a Rayleigh signal amplitude distribution, while initial phase is uniform. Here, instantaneous signal voltage changes have normal distribution with zero expected value and dispersion (at matched filter output) equalling $\overline{u_c^2}$.

Since noise at matched filter output has normal distribution with zero expected value and dispersion $U_{\text{ш}}^2$, then signal-plus-noise will have normal distribution with dispersion

$$U_{\text{сш}}^2 = \overline{u_{\text{c вых}}^2} + U_{\text{ш}}^2. \tag{9.73}$$

At moment t = T, filter output voltage envelope $u_{\text{вых}}(T)$ is compared with threshold $U_{\text{п}}$ (it is possible, without sacrificing conformity of results, to assume that the detector transfer constant equals unity). Therefore, error probabilities $P_{\text{л т}}$ and $P_{\text{пр}}$ completely are determined by the voltage $u_{\text{вых}}(T)$ law of distribution (uniform) and by the threshold $U_{\text{п}}$ magnitude.

It follows from what has been stated that the law of distribution of instantaneous filter output voltage changes will remain normal, both when there is and when is not a signal. The only thing that changes is the dispersion of this voltage from magnitude $U_{\text{ш}}^2$ to $U_{\text{сш}}^2 = U_{\text{ш}}^2 + \overline{u_{\text{c вых}}^2}$. Therefore, the envelope $u_{\text{вых}}(T)$ law of distribution must be a Rayleigh distribution when a signal is absent as well as when one is present. Consequently,

$$P_{\text{л т}} = \int_{U_{\text{п}}}^{\infty} \frac{u_{\text{ш вых}}(T)}{U_{\text{ш}}^2} e^{-u_{\text{ш вых}}^2(T)/2U_{\text{ш}}^2} du_{\text{ш вых}}(T) \tag{9.74a}$$

and

$$P_{\text{по}} = 1 - P_{\text{пр}} = \int\limits_{U_{\text{пор}}}^{\infty} \frac{u_{\text{сш вых}}(T)}{U^2_{\text{сш}}}\, e^{-u^2_{\text{сш вых}}(T)/2U^2_{\text{сш}}}\, du_{\text{сш вых}}(T), \tag{9.74b}$$

where $u_{\text{ш вых}}(T)$ and $u_{\text{сш вых}}(T)$ -- output voltage envelope values (at moment t = T) when a signal is absent and present, respectively.

As a result of integration, we obtain

$$P_{\text{лт}} = e^{-\frac{1}{2}\left(\frac{U_{\text{пор}}}{U_{\text{ш}}}\right)^2} \tag{9.75}$$

and

$$1 - P_{\text{пр}} = e^{-\frac{1}{2}\left(\frac{U_{\text{пор}}}{U_{\text{сш}}}\right)^2}. \tag{9.76}$$

It follows from (9.73) that

$$U^2_{\text{сш}} = U^2_{\text{ш}}\left(1 + \frac{\overline{u^2_{\text{с вых}}}}{U^2_{\text{ш}}}\right).$$

Since comparison with the threshold always will occur at moment t = T, then /137 $u_{\text{с вых}}$ should be understood to be $u_{\text{с вых}}(T)$, i. e.,

$$U^2_{\text{сш}} = U^2_{\text{ш}}\left[1 + \frac{\overline{u^2_{\text{с вых}}(T)}}{U^2_{\text{ш}}}\right]. \tag{9.77}$$

But, it follows from Chapter 2 that

$$\frac{u^2_{\text{c вых}}(T)}{U^2_{\text{m}}}=\frac{2Q}{N_0}=\frac{a^2T}{N_0}.$$

Averaging (with respect to realizations) the left and right sides of this equality, we have

$$\frac{\overline{u^2_{\text{c вых}}(T)}}{U^2_{\text{m}}}=\frac{\overline{a^2}T}{N_0}=\frac{Q_{\text{cp}}}{N_0}.$$

Substituting this expression into (9.77) and considering relationship (9.72), we obtain

$$U^2_{\text{cm}}=U^2_{\text{m}}(1+q). \tag{9.78}$$

Consequently, formula (9.76) may be written in the form

$$1-P_{\text{np}}=\exp\left[-\frac{1}{2}\left(\frac{U_{\text{op}}}{U_{\text{m}}}\right)^2\frac{1}{(1+q)}\right]. \tag{9.79}$$

Using formulas (9.71, (9.75), and (9.79), it is not difficult to compute error probabilities $P_{\text{лт}}$ and P_{np} for given values q and P(0)/P(C). Composite error probability is determined by formula

$$P_{\text{om}}=P(0)P_{\text{лт}}+P(C)P_{\text{np}}. \tag{9.80}$$

For example, we will examine a case where P(C) = P(0). Here, considering the normality condition, we obtain

$$P(C)=P(0)=0{,}5, \tag{9.81}$$

and the aforementioned formulas take the following form:

where

$$\left.\begin{aligned} &P_{\text{лт}}=e^{-\xi^2}, \quad P_{\text{пр}}=1-e^{-\xi^2/(1+q)},\\ &\xi^2=(1+1/q)\ln(1+q).\\ &P_{\text{ош}}=(P_{\text{лт}}+P_{\text{пр}})/2. \end{aligned}\right\} \tag{9.82}$$

The Figure 9.8 curve is plotted from these formulas.

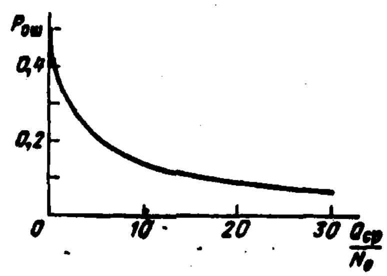


Figure 9.8

Given permissible probability $P_{\text{ош}}$ magnitude, it is possible to determine requisite signal-to-noise ratio q. If a priori probabilities P(0) and P(C) are unknown, composite error probability $P_{\text{ош}}$ also may not be found and it is necessary to be given conditional error probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$, rather than /138 magnitude $P_{\text{ош}}$. Here, as follows from (9.75) and (9.79), the requisite signal-to-noise ratio equals

$$q=\frac{\ln\dfrac{1}{P_{\text{лт}}}}{\ln\dfrac{1}{1-P_{\text{пр}}}}-1. \tag{9.83}$$

Where $P_{\text{пр}}\leqslant 0.1$, it is possible with sufficient precision to assume that

$$q\approx\frac{1}{P_{\text{пр}}}\ln\frac{1}{P_{\text{лт}}}. \tag{9.84}$$

Requisite threshold U_{op} magnitude is determined from relationship (9.75) and equals

$$U_{0p} = U_{ш} \sqrt{2\ln\frac{1}{P_{лт}}}. \qquad (9.85)$$

We will compare (9.84) with expression (9.49) obtained in § 9.2 for a signal with a precisely-known amplitude. In the case of a signal with precisely-known amplitude, requisite energy Q_0 will depend on error probabilities $P_{лт}$ and $P_{пр}$ in approximately an identical manner (where $P_{лт} \leqslant 0{,}1$ and $P_{пр} \leqslant 0{,}1$), and will do so logarithmetically. For a fluctuating signal, the dependence of average signal energy Q_{cp} on probabilities $P_{лт}$ and $P_{пр}$ is considerably different: dependence of Q_{cp} on $P_{лт}$ is logarithmetic, while it is hyperbolic on $P_{пр}$, i. e., significantly more defined. Therefore, for a fluctuating signal of the type being examined (i. e., for amplitude fluctuations with respect to Rayleigh's law), it is significantly more difficult to insure a slight miss probability than a slight false-alarm probability. It also follows from this that, given a slight permissible miss probability, detection of a fluctuating signal requires that the signal have considerably more energy than is the case for detection of a signal with known amplitude.

Analysis of the impact of the ambiguity (fluctuation) of signal rf occupation phase and its amplitude on required signal energy performed in the preceding sections demonstrates that, given slight permissible detection error probabilities, phase ambiguity does not play a significant role, while amplitude ambiguity may require a very significant increase in its average energy.

9.4 Impact of Receiver Frequency Characteristic and Bandwidth Shape on Detection Quality /139

It follows from the Figure 9.5 circuit that the basic optimum detector element is an optimum linear filter matched with the anticipated signal. This signifies that, for noise $u_{ш}(t)$ in the form of white noise, filter transfer constant $K(j\omega)$ must satisfy relationship (2.30), i. e., be complex conjugate with anticipated signal spectrum $S(j\omega)$. However, it was noted in § 2.3 that, during detection of a sinusoidal radio pulse (like the one depicted in Figure 2.7 or 2.19), there

is no requirement for strict satisfaction of that condition and approximate results may be obtained through replacement of the optimum linear filter by a standard band-pass amplifier with bandwidth matched with the width of the signal spectrum.

More detailed examination of this problem presented in [5] for a signal with known amplitude and random initial phase made it possible to plot the Figure 9.9 curves. In this figure, $\Delta f \tau_{и}$ — product of amplifier bandwidth Δf and signal radio pulse duration $\tau_{и}$ (with sinusoidal rf occupation). It is assumed here that time T, devoted to detection, coincides with the pulse duration ($T = \tau_{и}$); η — signal energy utilization factor, i. e,

$$\eta = \frac{Q}{Q'},$$

where Q and Q' — signal energy values required to insure a given detection quality (given detection error probabilities) when an optimum filter and standard band-pass

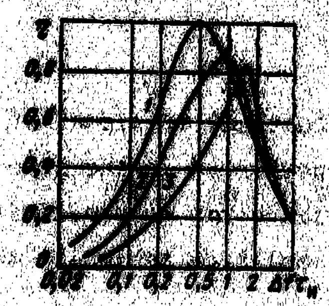


Figure 9.9

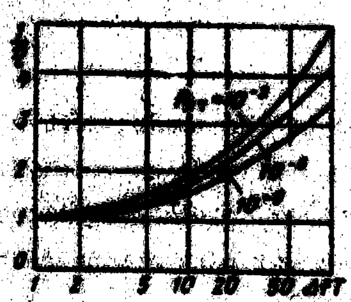


Figure 9.10

amplifier, respectively, are used. It is evident that magnitude $1/\eta$ equals requisite signal energy loss when a band-pass amplifier replaces an optimum filter. Curve 1 corresponds to a case where radio pulse envelope a(t) and band-pass amplifier frequency characteristic K(f) are described by gaussian curves. Curve 2 will relate to a case of identical frequency characteristic K(f), but for a pulse with a rectangular envelope. Curve 3 is valid for the case of rectangular a(t) and K(f).

It is evident from the figure that, given appropriate filter bandwidth Δf selection, the loss in requisite energy either is absent altogether (curve 1) or is very slight (curves 2 and 3), i. e., does not exceed 1 dB. In a case where there are slight band deviations from optimum value $\Delta f_{\text{опт}}$ (in particular, from value $1.37/\tau_{и}$ for curve 3), such as for changes not greater than a factor of 1.5, the loss rises slightly. For great deviations (for instance, for band broadening or narrowing by a factor of 5 - 10 and more), the loss is great.

However, when several additional conditions are met, a change in filter bandwidth by a factor of 5 - 10 or more may turn out to be permissible. We initially will examine these conditions relative to a case of broadening band Δf compared /140 to its optimum value $\Delta f_{\text{опт}}$. In this case, it suffices to connect an af filter with bandwidth ΔF matched with signal pulse duration at envelope detector output (Figure 9.5), i. e., bandwidth

$$\Delta F \approx \frac{1}{2\tau_{и}} = \frac{1}{2T}, \tag{9.86}$$

and, at moment t = T, compare it with this filter's output voltage threshold [it is evident here that the assumption is that the moment of anticipated appearance (or non-appearance) of a pulse at detector input and pulse duration $\tau_{и} = T$ are known beforehand at the point of reception].

We will examine a case where sinusoidal signal amplitude a is constant and known within observation cycle limits 0 - T, while initial phase is constant and equally probable in the interval $0 - 2\pi$. We will assume that $\Delta f \gg 1/\tau_{и}$ and, consequently, $\Delta f \gg \Delta F$. In this case, it is possible approximately to consider that af filter output voltage has a normal law of distribution, both when a signal is and is not present. Therefore, it is possible to assume that

$$\left.\begin{aligned} P_{\text{лт}} &= \frac{1}{\sqrt{2\pi}\,U_{\text{ш}}} \int_{U_0}^{\infty} \exp\left[-\frac{(U_{\text{ш}} - u_{\text{ш}})^2}{2U_{\text{ш}}^2}\right] du_{\text{ш}}; \\ 1 - P_{\text{пр}} &= \frac{1}{\sqrt{2\pi}\,U_{\text{сш}}} \int_{U_0}^{\infty} \exp\left[-\frac{(U_{\text{сш}} - u_{\text{сш}})^2}{2U_{\text{сш}}^2}\right] du_{\text{сш}}, \end{aligned}\right| \tag{9.87}$$

where $U_{=ш}$ and $U_{ш}^{2}$ -- noise constant component and dispersion at detector output when there is no signal, while $U_{=сш}$ and $U_{сш}^{2}$ -- analogous parameters when there is a signal (i. e., signal-plus-noise).

Dispersions $U_{ш}^{2}$ and $U_{сш}^{2}$ are determined within af filter bandwidth limits. This filter's transfer constant at the zero frequency is assumed to be unity. Here, since condition (9.86) is met, it is possible approximately to assume /141 that, by moment t = T of comparison with the threshold, average filter output voltage values and the dispersions of these voltages essentially achieve steady-state values.

Considering the aforementioned assumptions and also assuming $P_{пр} \leqslant 0{,}1$ and $P_{лт} \leqslant 0{,}1$, it is possible based on (9.87) to obtain the following expression valid both for a square and for a linear envelope detector:

$$q' = \frac{Q'}{N_0} \approx \sqrt{2\Delta f T}\left(\sqrt{\ln\frac{1}{P_{лт}}} + \sqrt{\ln\frac{1}{P_{пр}}}\right).$$

Comparing it with expression (9.49) for an optimum linear filter, we find out that

$$\frac{1}{\eta} = \frac{Q'}{Q} = \frac{\sqrt{2\Delta f T}}{\sqrt{\ln\frac{1}{P_{лт}}} + \sqrt{\ln\frac{1}{P_{пр}}}}. \qquad (9.88)$$

The result obtained in [127] for a fluctuating signal and $P_{лт}$ $P_{пр}$ coincides with formula (9.88) where ΔfT $\gg$ 1 (if in it one assumes that $P_{лт} \ll P_{пр}$), while, for ΔfT $\gtrsim$ 1 (and $P_{лт}$ $P_{пр}$)), it is depicted by the Figure 9.10 curves. The curves demonstrate that formula (9.88) is sufficiently precise already where ΔfT $\gtrsim$ 50 ÷ 100 (here, as was the case above, we assume that T = $\tau_{и}$). It also follows from this that, even when band Δf is broadened in comparison with $\Delta f_{опт} \approx 1/\tau_{и}$ by a factor of 5 — 10, the loss in requisite signal energy does not exceed 2 — 3 dB, i. e., is considerably less than when there is no af filter at detector output (compare the Figure 9.9 and 9.10 curves). Consequently, connection at detector output of an af filter with a band determined by relationship (9.86) makes the detector less critical to broadening of the bandwidth of the filter preceding the detector.

We now will explain under what conditions it is possible to insure slight detector criticality to a significant narrowing of the bandwidth of the filter preceding the detector (compared with $\Delta f_{опт} \approx 1/\tau_и$). As was the case previously, we will assume that the signal has the form of a sinusoidal pulse with a rectangular envelope and the moment of anticipated appearance (or non-appearance) of the signal at input is known. Also known is signal pulse duration $\tau_и$, equalling time T devoted to signal detection.

Under these conditions gating of filter $O\Phi$ preceding the detector also /142 is possible (Figure 9.5), i. e., connecting it only at time interval 0 ÷ T of

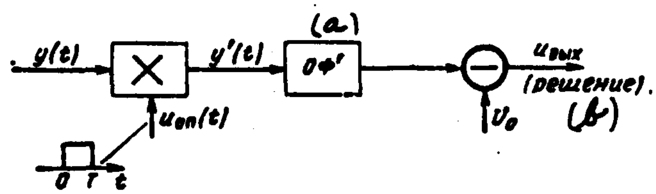


Figure 9.11. (a) -- OF' [optimum filter]; (b) -- Decision.

anticipated signal arrival. It is possible to consider the gating operation to be connection ahead of the filter (or beyond it) of a stage multiplying duration T by rectangular video pulse $u_{оп}(t)$, as shown in Figure 9.11.

Realized here in the optimum detector, in essence, is the optimum correlation-filtration device described in § 4.3 (see Figure 4.4). Gating voltage corresponds to function $f_1(t)$, while filter $O\Phi'$ must be matched with signal

$$u'_x(t) = f_2(t),$$

where $f_2(t)$ -- any function meeting condition (4.19), i. e., such that, following multiplication by $f_1(t)$, it provides an oscillation coinciding with anticipated signal $u_x(t)$.

It is clear from Figure 9.12 that it is possible to accept as function $u_x'(t)$ not only actual anticipated signal $u_x(t)$, but also any other signal of greater

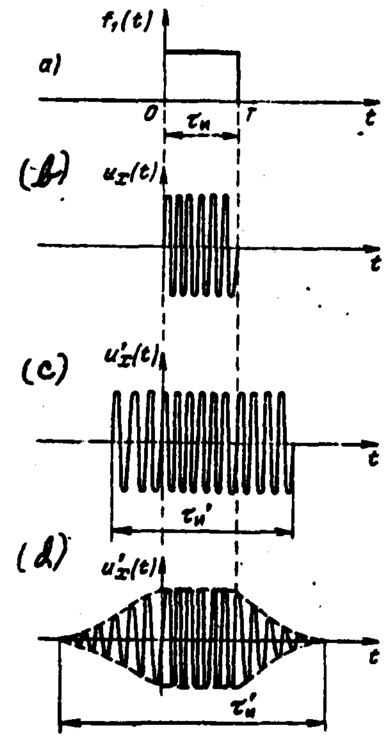


Figure 9.12

(but finite) duration which coincides with the anticipated signal only in sector 0 — T (the signal depicted in Figure 9.12c and 9.12d, for instance).

This denotes that, given the gating stage in the Figure 9.11 circuit, it is possible to select filter $O\Phi'$ matched with a signal of greater (than τ_u) duration and having any envelope shape outside the range 0 — T. But, optimum filter bandwidth narrows when signal duration increases, while the shape of this filter's frequency characteristic also changes with a change in envelope shape. Therefore, it is evident that, when gating is used, the impact of filter bandwidth narrowing (compared with $\Delta f_{onT} \approx 1/\tau_u$) and a change in frequency characteristic shape on detection quality radically decreases.

In addition, it follows from this that, if a quasi-optimum standard resonant amplifier with band Δf, for example, replaces the optimum filter, the filter bandwidth narrowing compared to $1/\tau_u$ also will have little impact on detection quality.

This result may be explained in the following manner. When gating is /143 used, there is a rise in both signal amplitude and noise dispersion in time interval $0 - T$ at filter output. If $\Delta f \ll 1/\tau_{\pi}$, then, by moment $t = T$, both signal voltage magnitude and noise dispersion magnitude turn out to be slight (compared to the steady-state value); therefore, there is no deterioration in the signal-to-noise ratio (in a real circuit, it is inadvisible all the same to select band Δf that is too slight since, here, the filter transfer constant will decrease radically and additional amplification must be used to compensate for this effect).

It follows from what has been stated that, in several cases, it is possible to obtain a detector slightly critical to significant (by several factors) changes in the bandwidth of the filter preceding the detector and to changes in the shape of this filter's frequency characteristic.

9.5 Binary Phase Detection

It was assumed during examination of binary detection in Chapter 5 and in the preceding sections of this chapter that a detector (receiver) optimally considers all usable information contained in signal-plus-noise.

It is convenient for relatively narrow-band signals to represent signal-plus-noise in the form of a single oscillation with some resultant amplitude $U_p(t)$ and resultant initial phase $\theta(t)$

$$y(t) = U_p(t) \cos[\omega t + \theta(t)]. \tag{9.89}$$

In the general case, information on the usable signal will be contained both in envelope $U_p(t)$ and in phase $\theta(t)$ of resultant oscillation $y(t)$. However, in some cases, usable information may be contained wholly (or almost wholly) only in envelope $U_p(t)$ or only in phase $\theta(t)$.

Actually, the above comparison of binary detection of precisely-known signals and of random initial phase signals demonstrated that, given slight signal-to-noise ratio, the uncertainty of the initial phase (or failure to take account of this phase in the detector) will lead to a large loss in requisite signal energy.

This loss decreases monotonically by virture of increasing the signal-to-noise ratio and, given sufficiently-strong signals, essentially equals zero.

This denotes that, in the case of a precisely-known signal, usable information will be contained both in envelope $U_p(t)$ and in phase $\theta(t)$ of the resultant oscillation. But, given a slight signal-to-noise ratio, this information is concentrated mainly in phase $\theta(t)$ (since, in this case, failure to account for $\theta(t)$, occurring in the amplitude detector, will lead to a sharp decrease in detection reliability), while it is concentrated mainly in envelope $U_p(t)$, when the signal-to-noise ratio is high.

Consequently, given slight slight signal-to-noise ratios, phase detectors, i. e., devices reacting only to phase $\theta(t)$ of the resultant oscillation and not reacting to its amplitude, may provide just as good results as an optimum /144 detector, which analyzes entire oscillation y(t). Therefore, along with the aforementioned detectors based on analysis of entire oscillation y(t) or its envelope $U_p(t)$, phase detectors are of great interest.

Interest in phase detectors especially rose when the following difficulty, which arises during practical realization of detectors which analyze instantaneous values y(t) and amplitude of oscillation $U_p(t)$, was encountered.

Threshold bias U_{op} is set at the output of such detectors. If a priori probabilities P(O) and P(C) are known, then requisite optimum magnitude U_{op} will depend both on signal-plus-noise input intensities, and on receiver amplification [see formulas (5.15) and (9.71), for example]. If probabilities P(O) and P(C) are unknown, then the threshold magnitude is established based only on the permissible false-alarm probability [see formulas (5.33b) and (9.85), for instance]. Here, requisite magnitude U_{op} will not depend on signal intensity, but dependence on input noise intensity and receiver amplification is retained. Therefore, in all cases, a change in receiver amplification (as a result of instability) requires corresponding adjustment in threshold U_{op} magnitude.

There is no such difficulty in phase detectors since phase $\theta(t)$ of the resultant oscillation will not depend on receiver amplification. In addition, in recent years radiotechnical phase systems, which include an amplitude detector,

have become widespread (see [177], for example). Here, it is important to know whether or not it is permissible to use limiter output voltage both for signal detection and for signal parameter measurement. For these reasons, a series of works [16, 124, 131, 177, 181, and others] has been devoted in recent years to examination of phase detection methods. We will dwell briefly on the most important results obtained in them.

Initially, we will examine <u>optimum</u> phase detection, in which information included in phase $\theta(t)$ is used in the best theoretically-possible manner. This denotes that the signal/no signal decision will be based upon comparison of inverse probabilities $P_\theta(C)$ and $P_\theta(0)$ of signal presence and absence for given realization $\theta(t)$ of input signal-plus-noise phase.

We will assume that signal $u_c(t)$ is precisely known, i. e.,

$$u_c(t) = a(t)\cos[\omega t + \varphi(t)], \tag{9.90}$$

where parameters a(t), $\varphi(t)$, and ω are precisely known. Then

$$\left.\begin{aligned} P_\theta(C) = kP(C)\,P_c(\theta), \\ P_\theta(0) = kP(0)\,P_0(\theta), \end{aligned}\right\} \tag{9.91}$$

where P(C) and P(0) -- signal/no signal a priori probabilities, while $P_C(\theta)$ and $P_0(\theta)$ -- multidimensional phase θ distributions for signal presence and absence, respectively.

A "yes" decision (signal) results in the optimum detector if /145

$$P_\theta(C) > \eta P_\theta(0), \tag{9.92}$$

and a "no" decision (no signal) results in the opposite case. Here, η -- weight factor considering the relative danger of signal false alarms and misses. If both types of errors are equally dangerous (i. e., there is a requirement to provide minimum composite error detection probability P_{om}), then

$$\eta = 1 \tag{9.93}$$

It follows from (9.91) and (9.92) that a "yes" decision is required if

$$\ln \frac{P_c(0)}{P_0(0)} > C, \tag{9.94a}$$

where

$$C = \ln\left[\eta \frac{P(0)}{P(C)}\right], \tag{9.94b}$$

and a "no" decision is required in the opposite case.

As was demonstrated in § 1.3, any time function f(t) with a relatively narrow-band spectrum with width Δf completely is characterized by n values (f_1, . . ., f_n) of this function or n_1 values of its envelope and n_1 values of phase [see (1.13) and (1.14), for example], where

$$n_1 = \frac{n}{2} = \Delta f \cdot T. \tag{9.95}$$

Here T — observation cycle duration.

In our case, signal-plus-noise y(t) plays the role of f(t), while its initial phase $\theta(t)$ plays the role of phase. Consequently, function θ (t) of interest to us completely is characterized by its selected n_1 values

$$(\theta_1, \theta_2, \ldots, \theta_{n_1}),$$

where n_1 is determined from formula (9.95).

Therefore,

$$\frac{P_c(0)}{P_0(0)} = \frac{P_c(\theta_1, \ldots, \theta_{n_1})}{P_0(\theta_1, \ldots, \theta_{n_1})}. \tag{9.96}$$

It is assumed further for simplicity that the values selected ($\theta_1, \ldots, \theta_{n1}$) statistically are independent. Then, from (9.94a) and (9.96) we obtain the following decision "yes" rule:

$$\ln \frac{\prod_{l=1}^{n_1} w_c(\theta_l)}{\prod_{l=1}^{n_1} w_0(\theta_l)} > C. \tag{9.97}$$

where $w_c(\theta_l)$ and $w_0(\theta_l)$ -- unidimensional probability densities of signal- /146 plus-noise initial phase θ and of noise alone, respectively, which are computed relatively simply (see [177], for instance). However, in spite of this, resultant expression (9.97) is unwieldly. Therefore, cases of a slight $(a/U_{m\,вх} \ll 1)$ and a high $(a/U_{m\,вх} > 1)$ signal-to-noise amplitude ratio at input usually are examined separately.

As already noted above, for a high signal-to-noise ratio, optimum phase detection must supply a large loss compared with completely-optimum detection [i. e, with detection based on processing of entire input realization y(t), rather than only its phase $\theta(t)$]. Mathematical calculations actually confirm this fact [131, 177, and others]. Therefore, we will dwell only on the case of slight signal-to-noise ratio $a/U_{m\,вх}$, for which optimum phase detection approximates the fully optimum.

At first glance, it would appear that a case of slight signal-to-noise ratio $a/U_{m\,вх}$ is not of interest since, here, it is impossible to insure high detection validity. However, in actuality, this is not the case, for detection error probabilities are determined by signal-to-noise power ratio Q/N_0 rather than amplitude ratio $a/U_{m\,вх}$. Therefore, if time T devoted to detection is sufficiently great,

then, even given slight ratio $a/U_{m\,sf}$, it is possible to obtain high power ratio Q/N_0 and, consequently, slight detection error probabilities.

Hence, it follows that, when phase detection is used, good results only may be obtained when the following two conditions are met simultaneously:

$$\frac{a}{U_{m\,sf}} \ll 1, \tag{9.98}$$

$$\frac{Q}{N_0} \gg 1, \tag{9.99}$$

where

$$U_{m\,sf} = N_0 \Delta f \tag{9.100}$$

-- extant value of the noise voltage contained in input signal-plus-noise y(t). Since $Q = a^2T/2$, then conditions (9.98) and (9.99) may be met simultaneously only where

$$\Delta f T \gg 1, \tag{9.101}$$

i. e., in cases where, during observation cycle T, one succeeds in obtaining a sufficiently-large number n_1 of independent selective values. When conditions (9.98) and (9.101) are met, analysis of decision rule (9.97) provides the following results (see [131] or [177], for example).

A signal present decision must be made if this inequality is satisfied

$$\int_0 a(t)\cos[\theta(t)-\varphi(t)]\,dt > U_0, \tag{9.102}$$

where $U_0 = \sqrt{\frac{N_0}{\Delta f}}\left(\sqrt{\frac{2}{\pi}}C + \sqrt{\frac{\pi}{8}}q\right).$ /147

$$q = \frac{Q}{N_0}. \tag{9.103}$$

If inequality (9.102) is not satisfied, a no signal decision results.

Hence, it follows that, under the aforementioned conditions $(a'U_{max} \ll 1,$ $\Delta fT \gg 1)$, an optimum phase detector must comprise an ideal phase detector*, which transforms the (9.89) input mixture into cos [θ (t) — ϕ (t)], a correlator, which computes the cross correlation of the phase detector output voltage and signal envelope a(t), and threshold bias U_0 at output. Here, it turns out that error probabilities $P_{\text{л т}}$ and $P_{\text{пр}}$ are determined by formulas differing from corresponding formulas (5.22) and (5.27b) for a completely-optimum detector only in that they will contain, in place of signal energy Q, magnitude

$$Q' = \frac{\pi}{4} Q. \tag{9.104}$$

This signifies that, when conditions (9.98) and (9.101) are met, an optimum phase detector provides identical detection error probabilities that a completely-optimum detector supplies when input signal energy is increased by a factor of $4/\pi$.

In other words, replacement of a completely-optimum detector by an optimum phase detector, given the aforementioned conditions, will lead to only a 1 dB loss in requisite signal energy magnitude. The structure of an optimum phase detector determined from algorithm (9.102) turns out to be more complex than that of a completely-optimum detector of the same signal (i. e., of a precisely-known signal). There is no difficulty in becoming convinced of this if you compare expressions (5.110 with (9.46b) and (9.102).

However, it is possible to demonstrate (see [177], for example) that, when conditions (9.98) and (9.101) are met, the Figure 9.13 phase detector may provide

---

*As opposed to a standard phase detector, ideal phase detector output voltage must not depend on input voltage envelope $U_p(t)$.

results essentially identical to the optimum phase detector. Here, Y -- wide-band amplifier; ИО -- ideal bilateral limiter. Amplifier bandwidth must be sufficiently broad $(\Delta f_{yc} \gg 1/T)$, to avoid an information loss concerning input signal-plus-noise y(t) phase $\theta(t)$.

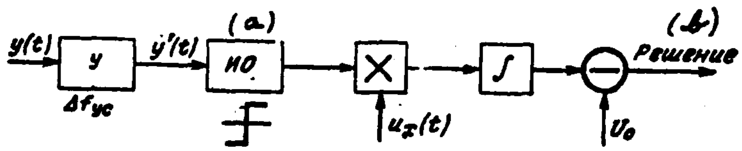


Figure 9.13. (a) -- IO [ideal limiter]; (b) -- D [decision].

Comparing the Figure 9.13 circuit with the Figure 5.4 completely-optimum /146 detector and considering that the correlator may be built based on the Figure 4.3 circuit, it is not difficult to become convinced that the given phase detector circuit, in essence, differs only by inclusion of the ideal limiter.

Inclusion of the limiter complicates the detector, but imparts to it the basic property inherent in a phase detector--insensitivity to amplification instability.

The results presented in this section relate to a precisely-known signal. It is possible in principle to use the phase approach also for detection of a random initial phase signal. However, mathematical calculations and detector structure significantly are complicated here (see [131 and [177], for example).

CHAPTER TEN /148

COMPLEX BINARY DETECTION

10.1 General Relationships

Let the signal have one of m values:

$$u_1(t),\ u_2(t),\ \ldots,\ u_m(t).$$

The requirement is to detect whether such a signal is present at system input or there is no signal, only noise.

Solution of this problem is equivalent to finding the answer to the question of whether any of m non-zero signals u_1, u_2, . . ., u_m (it is immaterial exactly which one) is present at system input or whether there is no signal, only noise. This task is called complex binary detection.

Detection is binary because only two types of answer are possible--"yes" (one of m non-zero signals is present) and "no" (no signal, only noise, is present), as was the case for binary detection examined in the preceding chapter.

However, in this case, detection is more complex since the signal has not one, but m possible types.

Such detection is of interest, for example, in radar, when the requirement /149 is to determine whether there is an aircraft in some region of space consisting of m elementary sectors, not specifying in exactly which sector it is located (it is assumed that only one aircraft may simultaneously be located in the entire region). Here, a return from an aircraft located in the k-th elementary sector may be considered possible signal $u_k(t)$ and it may be considered that establishment of the presence of one of possible signals $u_1, \ldots, u_m$ is equivalent to detection of the presence of an aircraft in the region being examined.

Let a priori probabilities $P(0), P(u_1), \ldots, P(u_m)$ of the absence and presence of signals $u_1, u_2, \ldots, u_m$, respectively, be known. Then, it is possible to compute corresponding inverse probabilities $P_y(0), P_y(u_1), \ldots, P_y(u_m)$.

It follows from the normality condition that

$$\left.\begin{aligned} P(0) + \sum_{k=1}^{m} P(u_k) = 1; \\ P_y(0) + \sum_{k=1}^{m} P_y(u_k) = 1. \end{aligned}\right| \tag{10.1}$$

Since appearances of signals $u_1(t), \ldots, u_m(t)$ are incompatible events, then inverse probability $P_y(C)$ of appearance of any of m possible signals (it is immaterial which one) equals the sum of the inverse probabilities of the appearance of each of these signals, i. e.,

$$P_y(C) = \sum_{k=1}^{m} P_y(u_k). \tag{10.2}$$

Analogously, for a priori probability P(C) of this event we have

$$P(C) = \sum_{k=1}^{m} P(u_k). \tag{10.3}$$

Solution of the signal detection problem using the maximum inverse probability

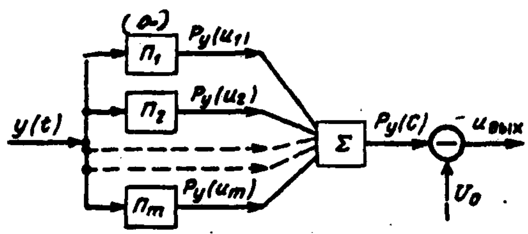


Figure 10.1

Key: a - P

approach requires computation and comparison between themselves of inverse probabilities $P_y(C)$ and $P_y(0)$ of the desired events. If it turns out that

$$P_y(C) > P_y(0), \tag{10.4}$$

then, the answer must be "yes" (one of m possible signals is present). In the opposite case, the answer is "no" (no signal, only noise). A receiver operating on this principle, as already noted in Chapter 4, provides minimum composite error decision probability. Here, no consideration is given to differences in the significance (danger) of various types of errors, i. e., of signal false alarms and misses. If a different "weight" is to be attached to false alarms and misses, then this may be considered by introduction of appropriate weight factor η into condition (10.4). Therefore, in the more general case, the decision /150 must be "yes" when this condition is met

$$P_y(C) > \eta P_y(0) \tag{10.5}$$

while the decision is "no" in the opposite case.

Considering (10.2), this condition takes the form

$$\sum_{k=1}^{m} P_y(u_k) > U_0, \tag{10.6}$$

where

$$U_0 = \eta P_y(0).$$

Consequently, the optimum receiving device must compute inverse probabilities $P_y(u_1), \ldots, P_y(u_m)$ of all possible signals, sum them, and compare them with some threshold U_0.

Therefore, the functional diagram of such a device in the general case must take the form depicted in Figure 10.1. In this figure, $\Pi_1, \Pi_2, \ldots, \Pi_m$ -- receivers computing inverse probabilities $P_y(u_1)$, $P_y(u_2)$, . . ., $P_y(u_m)$ of the corresponding signals.

In simple binary detection of some signal u_c, described in the preceding chapter, the job of the optimum receiver boils down to computation of inverse probability $P_y(u_c)$ of this signal and comparing it with some threshold. Therefore, each receiver Π_k in the Figure 10.1 system carries out functions identical (except for threshold comparison) as was the case for simple binary detection of the corresponding signal u_k and, consequently, has the same basic structure.

For our illustration, we will examine several specific cases of complex binary detection, assuming, as was the case previously, that noise is additive and white and has a normal distribution.

10.2 Detection of Precisely-Known Signals

Let all possible signals $u_1(t), \ldots, u_m(t)$ be precisely known. In this case, the task of each receiver Π_k boils down to computation of inverse probability $P_y(u_k)$ of a precisely-known signal.

It was found in Chapter 5 that /151

$$P_y(u_k) = k_2 P(u_k) e^{-Q_k/N_0} e^{\xi_k}, \tag{5.4}$$

where

$$\xi_k = \frac{2}{N_0} \int_0^T y(t) u_k(t) dt.$$

It was also demonstrated at that time that magnitude $P_y(u_k)$ may be computed by a receiver built according to the Figure 5.3 structural scheme. The basic element of such a receiver is correlator *KOP*, which, given several assumptions pointed out in Chapter 4 that usually are the case, may be replaced also by an optimum linear filter.

Consequently, in the case of precisely-known signals, each receiver Π_k in the Figure 10.1 schematic may be built according to the Figure 5.3 structure.

10.3 Detection of Random-Phase Signals

Let all signals $u_1(t)$, . . ., $u_m(t)$ be precisely known, with the exception of rf occupation phase, which is random and equally probable. In this case, the task of each receiver Π_k boils down to computation of inverse probability $P_y(u_k)$ of signal with a random equally-probable phase.

The appropriate expression for $P_y(u_k)$ was obtained in § 9.1 [see formula (9.15)] and may be written in the form

$$P_y(u_k) = k_3 P(a) e^{-Q_k/N_0} I_0\left(\frac{2a_k M_k}{N_0}\right). \tag{10.7}$$

where $Q_k = a_k^2 \tau_k/2$ -- signal $u_k(t)$ energy with amplitude a_k and duration τ_k.

Magnitude M_k is determined in accordance with formulas (9.12) and (9.13) in the following manner:

$$\left.\begin{aligned} M_k &= \sqrt{X_k^2 + Y_k^2},\\ X_k &= \int_{t_k}^{t_k+\tau_k} y(t)\cos\omega_k t\,dt;\\ Y_k &= \int_{t_k}^{t_k+\tau_k} y(t)\sin\omega_k t\,dt. \end{aligned}\right\} \tag{10.8}$$

where

Here, t_k and $t_k + \tau_k$ -- moment of signal $u_k(t)$ onset and disappearance, assumed to be known at the point of reception.

Formulas (10.7) and (10.8) are written for the general case where signals /152 $u_1(t)$, $u_2(t)$, . . ., $u_m(t)$ differ from one another with respect to amplitude a_k, carrier frequency ω_k, duration τ_k, and moment of onset t_k. If part of these parameters for all signals are identical, this simplifies formulas and optimum

system design accordingly. In the general case, the structural scheme has the form depicted in Figure 10.1, while each receiver Π_k accomplishes the operations required for computation of inverse probability $P_y(u_k)$ using formula (10.7).

As pointed out in § 9.1, magnitude M_k may be computed with the aid of electronic circuits depicted in Figure 9.1 or, given several limitations usually satisfied,

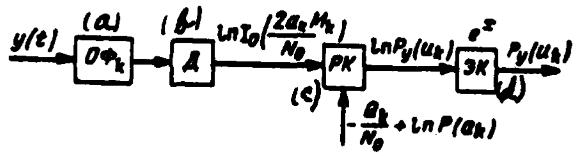


Figure 10.2. (a) -- OF_k [optimum filter k]; (b) -- D [detector]; (c) -- RK [difference stage]; (d) -- EK [exponential stage].

with the aid of an optimum linear filter. In the latter case, receiver Π_k may be built in accordance with the structural scheme depicted in Figure 10.2. Here, $O\Phi_k$ -- optimum linear filter matched with signal $u_k(t)$, while Д -- envelope detector separating the optimum filter output voltage envelope and clipping the value of this envelope at the moment signal $u_k(t)$ ceases, i. e., at moment $t_k + \tau_k$. Here, the response curve must be in the form $\ln I_0(x)$, rather than linear. As noted in § 9.1, such a curve may be obtained from a standard thermionic diode detector.

$-\ln P(a_k) + \frac{Q_k}{N_0}$ is computed in difference stage *PK* from voltage clipped by detector Д and equalling $\ln I_0\left(\frac{2a_k M_k}{N_0}\right)$. The result equals $\ln P_y(u_k)$ [since factor k_3 in formula (10.7) may be assumed to equal unity without harming the results]. Therefore, stage *ЭK* with an exponential characteristic converting $\ln P_y(u_k)$ into $P_y(u_k)$, is connected beyond the difference stage.

Since voltage corresponding to $P_y(u_k)$ is obtained at moment $t_k + \tau_k$, which can differ for different signals $u_k(t)$, then summing voltages of the type

$U_k = P_y(u_k)$, accomplished in stage Σ (Figure 10.1), in the general case also must include within itself the corresponding time delays of the individual summands.

10.4 Detection of Fluctuating Signals /153

Let all signals $u_1(t), \ldots, u_m(t)$ be fluctuating signals of the type examined in § 9.3.

In this case, every receiver Π_k must compute inverse probability $P_y(u_k)$, which, in accordance with formulas (9.63) and (.65), may be written in the form

$$P_y(u_k) = k_2 \frac{P(u_k)}{1 + \frac{Q_{cp\,k}}{N_0}} e^{h_k^2 M_k^2}, \tag{10.9}$$

where

$$h_k^2 = \frac{2\overline{u_k^2}}{N_0\left(1 + \frac{Q_{cp\,k}}{N_0}\right)}; \quad Q_{cp\,k} = \overline{u_k^2}\tau_k = \frac{\overline{a_k^2}}{2}\tau_k. \tag{10.10}$$

Magnitude M_k is determined exactly as it was in the preceding case. Factor k_2 during circuit design may be assumed to equal unity. Therefore, the receiver

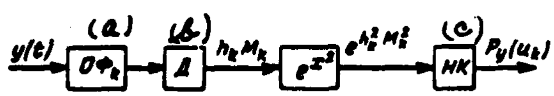


Figure 10.3. (a) -- OF_k [optimum filter k]; (b) -- D [detector]; (c) -- NK [normalization stage].

Π_k structural schematic in the examined case may be represented in the form depicted in Figure 10.3. Optimum linear filter $O\Phi_k$ has the same structure as in the Figure 10.2 circuit. Linear envelope detector Д, with transfer constant equalling h_k at moment $t_k + \tau_k$ clips voltage equalling $h_k M_k$. The stage with characteristic

$e^{r'}$ forms magnitude $e^{h_k^2 M_k^2}$, which, following the corresponding normalization in stage *HK*, provides desired inverse probability $P_y(u_k)$.

10.5 Detection of m Orthogonal Signals

Let signals $u_1(t), \ldots, u_m(t)$ be equally probable, differing only in position over time and thereby not overlapping over time.*

In this case, the general structural schematic of the optimum detection system (Figure 10.1) radically is simplified since, instead of m receivers of the Π_k type, a total of only one such receiver suffices.

Actually, since signals $u_k(t)$ are equally probable and differ only with /154 respect to possible time of arrival, it follows that inverse probability $P_y(u_k)$ of each signal may be computed with the aid of the identical circuit elements

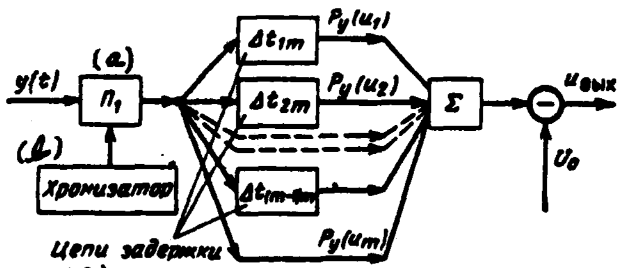


Figure 10.4. (a) -- P_1 [receiver 1]; (b) -- Timer; (c) -- Delay circuit.

without mode restructuring. Since possible signal $u_1(t), \ldots, u_m(t)$ values do not overlap in time and receiver Π_k must be connected only during time k when the anticipated signal is active, then the same receiver Π_k may be used sequentially (over time) to compute the probabilities $P_y(u_1), \ldots, P_y(u_m)$ of all possible signal values. Here, the Figure 10.1 circuit takes the form depicted in Figure 10.4, while Figure 10.5 depicts the time diagram characterizing this circuit's action.

*If signals $u_1, \ldots, u_m$ have parasitic random parameters, then it is assumed that a priori laws of distribution of these parameters are identical for all signals.

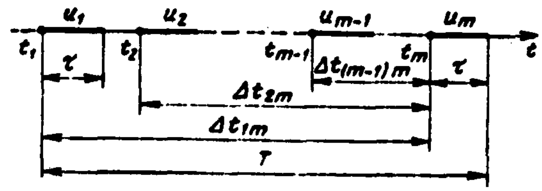


Figure 10.5

Possible signals $u_1(t)$, . . ., $u_m(t)$ have identical duration T and may arise only at the corresponding moments in time t_1, . . ., t_m known at the point of reception (Figure 10.5). At each moment in time, only one of these signals may exist or signals may be absent altogether, i. e., signal-plus-noise y(t) may contain only noise. The requirement is to determine whether one of these signals exists at system input or only noise exists. The timer (Figure 10.4) cuts in receiver $П_1$ sequentially for the intervals when anticipated signals are active to solve this problem (i. e., cuts it in at the intervals denoted by the heavy lines in Figure 10.5) and, by the beginning of each successive interval, insures disappearance of transient processes caused by oscillations arising in the receiver during the preceding interval. Then, receiver $П_1$ output voltage at moment $t_1 + \tau$ turns /155
out to equal $P_y(u_1)$, at moment $t_2 + \tau$, it equals $P_y(u_2)$, and so forth. Finally, at moment $t_m + \tau$, this voltage turns out to equal $P_y(u_m)$.

Therefore, for formation of the sum of the inverse probabilities included in inequality (10.6), voltages corresponding to values $P_y(u_1)$, . . ., $P_y(u_m)$, prior to summing in stage Σ, are passed beforehand across delay circuits, which cause delays at intervals Δt_{1m}, Δt_{2m}, and so forth, respectively.

Computation of detection error probabilities $P_{лт}$ and $P_{пр}$ requires finding the probability of satisfying inequality (10.6) when there is no signal and when there is a signal present, respectively. The left side of this inequality has a law of distribution differing from the norm and precise determination of this law in the general case presents great difficulties. Therefore, to date this problem only has been solved approximately and for the simplest cases.

The following cases were examined by Peterson and others [17]:

1. Signals $u_1(t), \ldots, u_m(t)$ are precisely known, equally probable, orthogonal, and have identical energies Q'.

In this case, energy Q', required to provide given error probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$ are determined from expression

$$\frac{Q'}{N_0} \approx \frac{1}{2} \ln\left[1 + m\left(e^{2Q/N_0} - 1\right)\right], \tag{10.11}$$

where Q -- energy of a precisely-known signal required in simple binary detection (i. e., where m = 1) to provide identical error probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$.

Where $P_{\text{пр}} \leqslant 0{,}1$ and $P_{\text{лт}} \leqslant 0{,}1$, magnitude Q/N_0, in accordance with formulas (5.31) and (5.31a) equals

$$\frac{Q}{N_0} \approx \left(\sqrt{\ln \frac{1}{P_{\text{лт}}} - 1{,}4} + \sqrt{\ln \frac{1}{P_{\text{пр}}} - 1{,}4}\right)^2. \tag{10.12}$$

2. Signals $u_k(t)$ have the form

$$u_k(t) = a_k(t) \cos(\omega t + \varphi),$$

where envelopes $a_k(t)$ and frequency ω are precisely known, while phase φ is random and equally probable within the range from 0 to 2π.

Signals $u_k(t)$ are equally probable and have identical energies Q', while envelopes $a_k(t)$ are orthogonal, i. e.,

$$\int_0^T a_k(t)\, a_l(t)\, dt = \begin{cases} 0 & \text{where } l \neq k, \\ 2Q' & \text{where } l = k. \end{cases} \tag{10.13}$$

Here, requisite energy Q' magnitude is determined from the equation

$$\ln\left[1 + \frac{1}{m} I_0\left(\frac{2Q'}{N_0}\right)\right] \approx \frac{2Q_0}{N_0}. \tag{10.14}$$

where Q_0 -- signal energy required in simple binary detection (i. e., where $m = 10$) and for identical error probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$. Relationships for determining energy Q_0 were given in § 9.2. Where $P_{\text{лт}} \leqslant 0{,}1$ and $P_{\text{пр}} \leqslant 0{,}1$, /156
magnitude Q_0/N_0 is determined from formula (9.49).

Relationships (10.11) and (10.14), obtained in [17], are approximate since, during their derivation, the true law of distribution of the log of the left side of inequality (10.6) was replaced by a normalized law, which is approximately valid only where $m \gg 1$.

For slight error probabilities ($P_{\text{лт}} \leqslant 0{,}1$ and $P_{\text{пр}} \leqslant 0{,}1$), formulas (10.11) and (10.14) may be simplified significantly.

Initially, we will examine formula (10.11), i. e., the case of a precisely-known signal. Where $m \gg 1$, $P_{\text{лт}} \leqslant 0{,}1$, and $P_{\text{пр}} \leqslant 0{,}1$, the result is $me^{2Q/N_0} \gg 1$ and from (10.11) we have

$$\frac{Q'}{N_0} \approx \frac{1}{2} \ln m + \frac{Q}{N_0}. \tag{10.15}$$

Considering relationship (10.12), finally we obtain

$$\frac{Q'}{N_0} = \left(\sqrt{\ln \frac{1}{P_{\text{лт}}} - 1{,}4} + \sqrt{\ln \frac{1}{P_{\text{пр}}} - 1{,}4}\right)^2 + 0{,}5 \ln m. \tag{10.16}$$

Now we will examine formula (10.14) for a random-phase signal.

When $P_{\text{лт}} \leqslant 0{,}1$ and $P_{\text{пр}} \leqslant 0{,}1$, the result is $2Q_0/N_0 \geqslant 12$; here, as follows from (10.14), it must be that

$$1 + \frac{1}{m} I_0\left(\frac{2Q'}{N_0}\right) > 10^6,$$

therefore, in formula (10.14), it is possible to assume with great precision:

$$\frac{1}{m} I_0\left(\frac{2Q'}{N_0}\right) \gg 1, \quad I_0\left(\frac{2Q'}{N_0}\right) \gg 1.$$

and

$$I_0\left(\frac{2Q'}{N_0}\right) \approx \frac{e^{2Q'/N_0}}{\sqrt{2\pi\frac{2Q'}{N_0}}}.$$

Considering these relationships, formula (101.4) is reduced to the form

$$\frac{Q'}{N_0} \approx \frac{Q_0}{N_0} + 0{,}5\ln m. \tag{10.17}$$

Since magnitude Q_o/N_o is determined by formula (9.49), then it is possible to write

$$\frac{Q'}{N_0} \approx \left(\sqrt{\ln\frac{1}{P_{\text{лт}}}} + \sqrt{\ln\frac{1}{P_{\text{пр}}} - 1{,}4}\right)^2 + 0{,}5\ln m. \tag{10.18}$$

As follows from what has been stated, formulas (10.15) and (10.17) [or (10.16) and (10.18)] were obtained with the assumption that $m \gg 1$. However, it follows from expressions (10.15) and (10.17) that they provide the correct result also where m = 1 ($Q' = Q$ and $Q' = Q_o$, respectively).

Comparison of formulas (10.16) and (10.18), valid for precisely-known /157 and for random-phase signals, respectively, demonstrates that, where $P_{\text{лт}} \leqslant 0{,}1$ and $P_{\text{пр}} \leqslant 0{,}1$, phase irregularity will lead to an insignificant (less than 2 db) increase in requisite signal energy. The less $P_{\text{лт}}$ and $P_{\text{пр}}$ and the greater m, the less the impact of signal phase irregularity.

CHAPTER ELEVEN /157

DETECTION AND RECOGNITION OF SIGNALS WITH MANY POSSIBLE VALUES

11.1 Optimum Receiver Structure

In Chapter 10, a case was examined in which the requirement is to detect whether or not any of m possible non-zero signals $u_1(t), \ldots, u_m(t)$ is present at input, without refinement of exactly which signal is present. In this chapter, we will examine a case where the requirement is not only to determine whether or not any of m non-zero signals is present (i. e., to solve the detection problem), but also to indicate exactly which of the possible signals is present (i. e., to solve the signal recognition* problem). Therefore, such a case may be referred to as signal detection with recognition. This occurs, for instance, during aircraft detection by a radar with a plan-position indicator. Actually, such a radar must determine whether or not there is an aircraft (or any other desired object) within an investigated region of space comprising m elementary sectors and, if there is, then in exactly which of these sectors it is located (it is assumed that only one aircraft may be located simultaneously in the entire region). Here, the return from an aircraft located in the k-th elementary sector may be considered possible

*Lately, this problem often is referred to as signal <u>classification</u>; the term <u>recognition</u> is used only for discrimination of complex models.

signal $u_k(t)$ and the radar task is equivalent to the multivalue signal detection and recognition task formulated above.

As indicated in Chapter 5, solution of such a problem with minimum composite error probability requires comparison between themselves of inverse probabilities $P_y(u_o)$, $P_y(u_1)$, . . ., $P_y(u_m)$ of all possible signals and selection of that signal whose inverse probability turns out to be greatest.

Here, a case of no signal (i. e., only noise at input) is considered /158 presence of signal u_o identical with zero, i. e.

$$u_0(t) \equiv 0. \tag{11.1}$$

All possible errors are considered equally dangerous in this solution to the problem. If it is desirable to attach different weights to different errors, then it is necessary to compare inverse probabilities taken with corresponding weights η_o, η_1, . . ., η_m, i. e., to select the greatest of the following magnitudes:

$$\eta_0 P_y(u_0), \; \eta_1 P_y(u_1), \; \ldots, \; \eta_m P_y(u_m).$$

However, for simplicity, in future all weights are assumed to be identical.

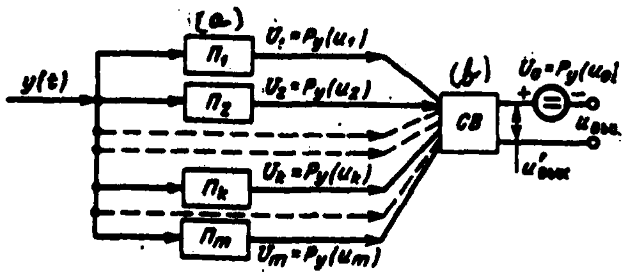


Figure 11.1. (a) -- P [receiver]; (b) -- SV [sampling circuit].

Then, the optimum receiver structural diagram may be represented in the form depicted in Figure 11.1.

In this figure, inverse probabilities $P_y(u_1)$, . . ., $P_y(u_m)$ are computed by receivers $\Pi_1, \ldots, \Pi_m$, while inverse probability $P_y(u_0)$ is considered some constant bias U_0. Sampling circuit *CB* selects from voltages U_1, . . ., U_m, equalling inverse probabilities $P_y(u_1)$, . . ., $P_y(u_m)$, the greatest voltage, designated $u'_{\text{вых}}$. Further, this voltage is compared with threshold bias U_0 and, if it turns out that

$$u'_{\text{вых}} \leqslant U_0, \text{ i. e., } u_{\text{вых}} \leqslant 0, \tag{11.2}$$

then a decision is made that no signals u_1, . . ., u_m are present.

In the opposite case, the response is that one of m non-zero signals is present at system input, this being that signal u_k to which greatest voltage U_k value corresponds.

Since the job of receiver Π_k is to compute inverse probability $P_y(u_k)$, i. e., solution of the same problem as was the case in the detector circuit (Figure 10.1), then their structure and operating mode may be selected so that receiver Π_k may be built in accordance with the Figure 10.2 circuit, or from the Figure 10.3 circuit in the case of a fluctuating signal. However, the receiver Π_k structure in Figure 11.1 may be simplifed since, here, any simultaneous monotonic /159
nonlinear conversion of voltages U_0, U_1, . . ., U_m is permissible.

Actually, in the Figure 11.1 circuit, any decision made is because each of the magnitudes U_0, U_1, . . ., U_m turns out to be the greatest. Therefore, if all these magnitudes are subjected to the identical inertia-free nonlinear transformation of the type

$$U'_k = f(U_k),$$

where $f(U_k)$ -- any monotonic function, and transformed magnitudes U_k' are compared among themselves, then the greatest magnitude U_k wi'' be transformed into the greatest magnitude U_k' and, consequently, the result of the comparison remains unchanged. Thus, for example, it is evident from Figure 11.2 that, if magnitude U_2 turns out to be the greatest, then voltage U_2' also will be the greatest.

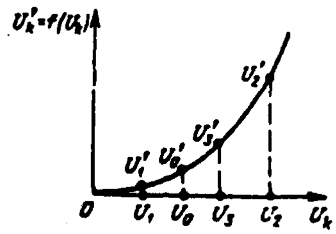


Figure 11.2

Consequently, it is permissible in the optimum detection with recognition system (Figure 11.1) to use any identical monotonic transformation of all output voltages $U_1, \ldots, U_m$, if threshold bias U_0 here also is subjected to the same transformation, i. e., to adjust the magnitude of this bias accordingly. Such system invariance to random monotonic output voltage transformations makes it possible in several cases to simplify receiver $П_k$ output stages considerably.

For example, we will examine the case of fluctuating signals. Here, in accordance with (10.9), we have

$$P_y(u_k) = k_3 \frac{P(u_k)}{1 + \frac{Q_{\text{гр}\,k}}{N_0}} e^{h_k^2 M_k^2}. \tag{10.9}$$

We will assume that all signals are equally probable and have identical energies Q. Then, it is possible to represent expression (10.9) in the form

$$U_k = P_y(u_k) = Ce^{h_k^2 M_k^2}. \tag{11.3}$$

where C — constant independent of the k-th signal. Here, the receiver $П_k$ circuit has the form depicted in Figure 10.3, but the normalizing stage is absent. Consequently, receiver $П_k$ may comprise optimum filter $ОФ_k$, linear amplitude detector $Д$, and a stage with a characteristic of the type e^{x^2}.

In accordance with what was stated above, it is possible to use any monotonic

transformation to magnitude $P_y(u_k)$(identical for numbers k). In this case, receiver structure simplification requires use of a transformation of the type

$$U_k' = \sqrt{\ln \frac{U_k}{C}}. \qquad (11.4)$$

This transformation is monotonic with respect to U_k and, consequently, is permissible. /160

From (11.3) and (11.4), we obtain

$$U_k' = h_k M_k. \qquad (11.5)$$

Consequently, instead of comparison among themselves and with threshold U_0 of type (11.3) expressions U_k, it is possible to compare much more simply expression

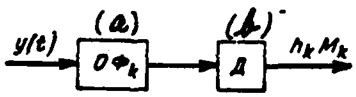


Figure 11.3

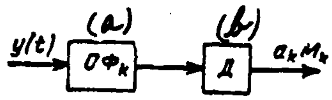


Figure 11.4

KEY: (a) -- OF_k [optimum filter k]; (b) -- D [detector].

(11.5), having changed the threshold here from U_0 to U_0'. This denotes that it is possible to remove the characteristic of the type e^{x^2} from the Figure 10.3 circuit and it will take on the form depicted in Figure 11.3. In this circuit, Д is a linear amplitude detector.

Now, we will examine a case where signals $u_k(t)$ have known amplitudes and random equally-probable phases. Here, receiver Π_k structure will have the form depicted in Figure 10.2. Again, we will assume that all signals u_k are equally probable and have identical energies. Then, difference stage *PK* in the Figure 10.2 circuit may be removed and receiver Π_k will comprise optimum filter $O\Phi_k$ and an envelope detector with response curve of the type $I_0\left(\frac{2a_k M_k}{N_0}\right)$. Since this curve

is monotonic with respect to magnitude $a_k M_k$ changes, then it is possible to compare voltages proportional to $a_k M_k$, instead of voltages proportional to $I_0\left(\frac{2a_k M_k}{N_0}\right)$, due to system invariance to monotonic transformations. This denotes that receiver Π_k in the Figure 11.1 system may be built in accordance with the Figure 11.4 layout. In this circuit, $Д$ is a linear amplitude detector. Here, threshold bias in the Figure 11.1 circuit must be changed accordingly (from U_0 to U_0').

Comparing the Figure 11.3 and 11.4 circuits, obtained for a fluctuating signal and for one with known amplitude, respectively, it is easy to become convinced that they differ only in factors h_k and a_k. If signals $u_k(t)$ have not only identical energies (as was accepted above), but identical duration as well, then factors h_k and a_k will not depend on number k and, therefore, may be omitted (given the corresponding adjustment of output threshold bias U_0). Here, the Figure 11.3 and 11.4 circuits coincide completely. This denotes that, given the aforementioned assumptions, receivers Π_k in the optimum detection with recognition system (Figure 11.1) turn out to be identical for fluctuating signals and for those with known amplitude. The only difference is threshold bias U_0 magnitudes. Consequently, the transition in the Figure 11.1 circuit from optimum reception of signals with known amplitude to optimum reception of fluctuating signals may be made /161 without any changes in the receiving system circuit, with the exception of the corresponding adjustment of output bias U_0.

Thus, with the assumptions made above, an optimum detection with recognition system has invariance, not only with respect to monotonic transformation of the system's response curve shape, but also to a change in the law of signal amplitude distribution as well. This property makes it possible to make the system simpler and more standard than is the case for optimum detection without recognition (Figure 10.1).

11.2 Detection and Recognition Error Probabilities

The following types of errors are possible in a signal detection with recognition system (Figure 11.1):

1. False alarms, i. e., responses concerning presence of some signal when,

in actuality, there is no signal at input. $P_{\text{л т}}$ denotes the probability of this error.

2. Distortions, i. e., incorrect responses in cases when there is some signal at input [signal misses or incorrect indication of the number k of present signal $u_k(t)$]. $P_{\text{иск}}$ designates the probability of such a distortion.

3. It follows from point 2 that a signal miss (i. e., a response concerning signal absence at input when, in actuality, any of m non-zero signals is present) is a particular case of distortion. Therefore, miss probability $P_{\text{пр}}$ always is less than or equal to distortion probability:

$$P_{\text{пр}} \leqslant P_{\text{иск}}. \tag{11.6}$$

Composite error probability $P_{\text{ош}}$ and composite correct response probability $P_{\text{прав}}$ corresponding to it may be computed from formulas (5.7) and (5.8). Along with composite probability $P_{\text{прав}}$, it often is convenient also to include conditional correct response probability $P_{\text{с прав}}$, determined for a condition where any one of the non-zero signals is present at input.

Evidently,

$$P_{\text{с прав}} = 1 - P_{\text{иск}}. \tag{11.7}$$

In the general case, error probability computation represents considerable difficulties. Therefore, we will limit ourselves to examination of a rather-particular, but important, case meeting the following conditions:

a) noise is additive normal white noise;

b) all non-zero signals $u_1(t)$, . . ., $u_m(t)$ are equally probable, orthogonal, and have identical energies.*

This case was examined in § 5.4 for precisely-known signals and in the /162

---

*If signals $u_1(t)$, . . ., $u_m(t)$ have parasitic random parameters, then the laws of distribution of these parameters for all signals are assumed to be identical.

assumption that the a priori zero signal probability equals zero. In this section, we examine a more-general case where signals may have random parameters, while the a priori zero signal probability does not equal zero.

As already noted in Chapter 5, signal orthogonality will lead to the fact that their distortion by noise turns out to be statistically independent. In

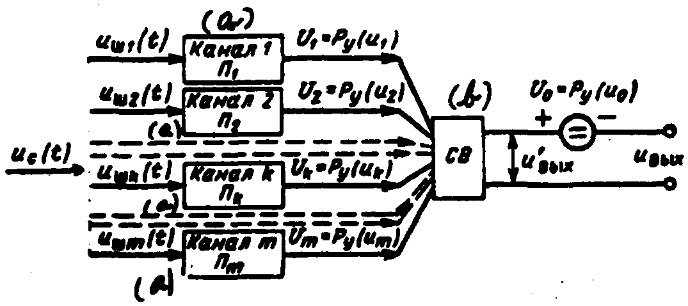


Figure 11.5. (a) -- Channel number, receiver number;
(b) --SV [sampling circuit].

addition, it also was shown that relationships obtained for such signals turn out to be valid also for the Figure 5.6 m-channel receiving device. In this case, considering that the signal also may have a zero value, the m-channel receiving device schematic takes the form depicted in Figure 11.5.

As before, we assume that additive noise voltages $u_{ш1}(t), \ldots, u_{шm}(t)$ active at channel input have normal laws of distribution with zero mean values and identical dispersions N (here, N = $\overline{u^2_{шk}}$). Signal $u_c(t)$ (precisely known or with parasitic random parameters) with equal probability may be present at input of any of m channels or be absent in all channels. The task of the receiving device in this case comprises determining whether signal $u_c(t)$ is present at input of any channel and, if so, at input of exactly which channel.

All Figure 11.5 designations correspond to the Figure 11.1 designations. Optimum receivers $П_1, \ldots, П_k$ in the Figure 11.5 schematic have a structure identical to that of the Figure 11.1 system, if the latter is designed for receipt of signals

not overlapping over time, i. e., if signal orthogonality is achieved because they do not overlap over time.

Probabilities $P_{лт}$ and $P_{пр}$ have the exact concept in the case of an m-channel system, as is the case of m possible signals.

The term distortion probability $P_{иск}$ in an m-channel system is understood to mean the probability of an incorrect response when signal $u_c(t)$ is present, i. e., a signal miss or incorrect indication of the number of the channel in which a signal is present. Here, formulas for determination of error probabilities $P_{лт}$, $P_{пр}$ and $P_{иск}$ in both cases (i. e., in the case of m orthogonal signals and in an m-channel system) are obtained in the identical way. Therefore, in future /163

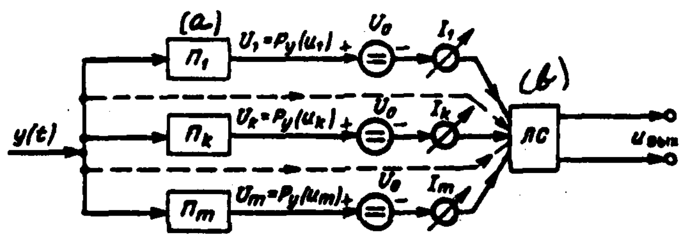


Figure 11.6. (a) -- P [receiver]; (b) -- LS [arithmetical unit].

for precision we will make our examination relative to a case of m orthogonal signals, i. e., to the Figure 11.1 system. Here, it becomes convenient for analysis to replace the optimum detection with recognition system (Figure 11.1) with the system depicted in Figure 11.6. This system in future is called quasi-optimum since probabilities $P_{лт}$ and $P_{пр}$ in it are obtained in exactly the same manner as in the optimum system, but distortion probability $P_{иск}$ is somewhat greater.

Receivers $П_1, \ldots, П_m$ and threshold bias U_0 in a quasi-optimum system are identical to those in the optimum system, but bias U_0 is connected at the output of each system receiver $П_k$ rather than at system output.

The threshold bias overruns of all channels are added in the ЛС section, i. e.,

$$u_{вых}=\sum_{k=1}^{m}\Delta U_k,$$

where

$$\left.\begin{aligned}&\Delta U_k=U_k-U_0 \quad \text{where} \quad U_k>U_0;\\ &\Delta U_k=0 \qquad\quad\ \ \text{where} \quad U_k\leqslant U_0.\end{aligned}\right\} \qquad (11.8)$$

Evidently, $u_{вых}\geqslant 0$ is always the case.

If $u_{вых}=0$, then it is considered that there is no signal at system input; if $u_{вых}>0$, then the decision is that there is a signal at input and it is that very signal $u_k(t)$ for which it turned out that $U_k > U_0$* (here, the arrow of the appropriate indicator I_k dips).

We will demonstrate that, in this principle of quasi-optimum system operation, false-alarm and miss probabilities arising in it will be identical to those in an optimum system, while distortion probability is somewhat greater.

In the optimum system (Figure 11.1), a false alarm will occur if $u'_{max}>U_0$, /164 is the case when there is no signal and will not occur if $u'_{max}\leqslant U_0$, where u'_{max} is the greatest of voltages $U_1, \ldots, U_m$.

But, if the greatest of voltages $U_1, \ldots, U_m$ exceeds U_0, then there will be a false alarm in the quasi-optimum system (Figure 11.6) as well; if the greatest of voltages $U_1, \ldots, U_m$ does not exceed U_0, then there will be no false alarm in the quasi-optimum system as well (Figure 11.6).

Consequently, cases when there are and are not false alarms in the systems examined (Figures 11.1 and 11.6) always coincide, which means that false-alarm probabilities $P_{лт}$ also coincide.

*If threshold U_0 is exceeded in more than one channel when there is a signal, then there is no signal number indication (i. e., recognition) and it is assumed that a distortion is occurring.

Analogously, one may show that miss probabilities $P_{пр}$ in both systems (Figures 11.1 and 11.6) also coincide.

We now will compare distortion probabilities $P_{иск}$.

It follows from (11.7) that, instead of distortion probabilities $P_{иск}$, it is possible to compare among themselves conditional probabilities $P_{c\ прав}$, i. e., probabilities of correct response for the condition where any one of m signals is present at input.

In the Figure 11.1 system, when signal number k is present, a correct response is dependent on voltage U_k at appropriate receiver $П_k$ output being greater than at the output of the remaining receivers and, in addition, it must exceed threshold U_o.

In the quasi-optimum system (Figure 11.6), when signal number k is present, a correct response depends not only on receiver $П_k$ output voltage U_k being greatest and exceeding threshold U_o, but also on the output voltages of all remaining receivers being less than U_o.

Consequently, the conditions for obtaining a correct response are stricter in the quasi-optimum system than in the optimum system and they will be met in fewer cases. Therefore, conditional correct response probability $P_{c\ прав}$ is less in a quasi-optimum system than in the optimum system.

Thus, if false-alarm, miss, and distortion probabilities in the optimum system equal $P_{л\ т}$, $P_{пр}$ and $P_{иск}$, then, in the quasi-optimum system, the probabilities of false alarms and misses will be identical, while distortion probability $P'_{иск}$ will be less, more precisely

$$P'_{иск} \geqslant P_{иск}. \qquad (11.9)$$

We will compute error probabilities in the quasi-optimum system (Figure 11.6). This system will comprise m channels, each comprising a receiver $П_k$, threshold bias U_o, and bias overrun indicator I_k.

We will introduce probabilities $P_{\text{лт}\,k}$ and $P_{\text{пр}\,k}$ for each such channel, having determined them in the following manner:

$P_{\text{лт}\,k}$ is the probability of bias U_0 overrun in the k-th channel, when there is no signal with this number at input;

$P_{\text{пр}\,k}$ is the probability that bias U_0 in the k-th channel will not be /165 exceeded when signal $u_k(t)$ with this number is at system input.

Initially, we will show that probabilities $P_{\text{лт}\,k}$ and $P_{\text{пр}\,k}$ determined in this manner, in the case of orthogonal signals, coincide with the false-alarm and miss probabilities, respectively, in each channel, given simple binary detection.

If the k-th channel is operating in the simple binary detection mode, then the false-alarm probability in this channel is the probability that threshold U_0 will be exceeded at its output when only noise is present at input. Probability $P_{\text{лт}\,k}$, as accepted above, is determined for a condition where a signal only of given number $u_k(t)$ is absent at the k-th channel input. This signifies that, given determination of magnitude $P_{\text{лт}\,k}$ at the input of a given numbered channel, a signal of any other number 1 may be present, where $1 \neq k$ (or there may be no signal at all). However, since all signals are orthogonal, they do not overlap with respect to time or to frequency spectra, for instance, then a signal only of given number k may pass through receiver Π_k, while signals of all remaining numbers exert no impact whatsoever on this receiver's output voltage (nor on the probability of a threshold U_0 overrun or non-overrun). Thus, for example, if signals do not overlap with respect to time, then each receiver Π_k opens only for a time interval from t_k to $t_k + \tau_k$, in which only anticipated signal $u_k(t)$ may be active. During action of any remaining anticipated signals, receiver Π_k will be blanked.

If signal orthogonality is achieved so that they do not overlap with respect to frequency spectra, then receivers Π_k have nonoverlapping bandwidths in all input networks and the spectra of all anticipated signals except $u_k(t)$ fail to reach the bandwidth of each receiver Π_k.

Thus, only noise and signal $u_k(t)$ may impact upon receiver Π_k output voltage, and signals $u_1(t)$ (where $1 \neq k$) exert no impact whatsoever. Hence, it follows that probability $P_{\text{лт}\,k}$ equals the false-alarm probability in the k-th channel,

given simplex signal $u_k(t)$ binary detection. It will not be difficult analogously to become convinced that probability P_{np} equals the signal miss probability in the k-th channel, given simple signal $u_k(t)$ binary detection.

False-alarm and miss probabilities for simple binary detection were found earlier in Chapter 9 and simple formulas were obtained for certain signal types. Thus, for example, formula (9.83) is valid for simple binary detection of a fluctuating signal and, in this case, may be written in the form

$$\frac{Q_{cp}}{N_0}=\frac{\ln\frac{1}{P_{\pi\tau k}}}{\ln\frac{1}{1-P_{np\,k}}}-1, \tag{11.10}$$

where Q_{cp} -- signal u_k ( ) average energy (since it was accepted that all /166 possible signals have identical energies, then Q_{cp} will not depend on number k).

Consequently, if the permissible error probabilities $P_{\pi\tau k}$ and $P_{np\,k}$ in each channel are known, then the corresponding simple binary detection formulas may be used to determine requisite signal energy. Therefore, it remains only to find the link between the error probabilities in each channel and in the system as a whole, i. e., the relationship of probabilities $P_{\pi\tau}$, P_{np} and $P_{\text{нсн}}$ to $P_{\pi\tau k}$ and $P_{np\,k}$. To do so, we again turn our attention to the Figure 11.6 diagram and we will consider that, in this case, when all signals are accepted as being equally probable with identical average energies and identical laws of parasitic random parameter distributions, probabilities $P_{\pi\tau k}$ and $P_{np\,k}$ in all channels are identical, i. e., will not depend on number k (in spite of this, we will retain index x in order to underscore that the given probabilities relate to one system channel, and not to the system as a whole).

Initially, we will find false-alarm probability $P_{\pi\tau}$ or, which is more convenient, probability $(1-P_{\pi\tau})$ of the absence of a false alarm in the system.

In order for there to be no false alarm in the examined system, there is a requirement that the threshold not be exceeded in any one of the channels when there is no signal, i. e., that there be no false alarms in any channel. But, the probability of the absence of a false alarm in a given channel equals

$(1-P_{\text{лт}k})$, while errors in all channels statistically are independent in light of the aforementioned assumptions; therefore

$$1-P_{\text{лт}}=(1-P_{\text{лт}k})^m \text{ and } P_{\text{лт}}=1-(1-P_{\text{лт}k})^m. \tag{11.11}$$

Now, we will find miss probability $P_{\text{пр}}$. Evidently

$$P_{\text{пр}}=P_0(u_1)P_{1\text{пр}}+P_0(u_2)P_{2\text{пр}}+\ldots+P_0(u_m)P_{m\,\text{пр}}, \tag{11.12}$$

where $P_0(u_1)$, . . ., $P_0(u_m)$ -- a priori probabilities of the presence of signals u_1, . . ., u_m, respectively, with the condition that any one of these signals must be present, i. e.,

$$P_0(u_1)+\ldots+P_0(u_m)=1. \tag{11.13}$$

The designation P_k is used for the signal miss probability in the system for a condition where signal $u_k(t)$ is present at its input (do not confuse conditional probability $P_{k\,\text{пр}}$, relating to the system as a whole, with probability $P_{\text{пр}\,k}$, relating to one of this system's channels).

In light of the assumptions made above concerning the nature of the signals, we obtain

$$P_{1\text{пр}}=P_{2\text{пр}}=\ldots=P_{m\,\text{пр}};$$

therefore, formula (11.12) provides

$$P_{\text{пр}}=P_{1\text{пр}}[P_0(u_1)+\ldots+P_0(u_m)].$$

Considering normality condition (11.13), we obtain

$$P_{\text{пр}}=P_{1\text{пр}}, \tag{11.14}$$

i. e., for the signals being examined, when computing miss probability $P_{\text{пр}}$ /167 in the system as a whole, one may assume that signal $u_1(t)$ is present at system input.

When signal $u_1(t)$ is present, a miss occurs in the system (Figure 11.6) if threshold U_0 is not exceeded in any one of the system's m channels. But, with respect to the determination, the probability that, given signal $u_1(t)$ presence, the threshold will not be exceeded in the first channel is $P_{\text{пр}\,1}$, meanwhile $P_{\text{пр}\,1} = P_{\text{пр}\,k}$. Given signal $u_1(t)$ presence, the probability that the threshold will not be exceeded in the k-th channel (where $k \neq 1$) is $1 - P_{\text{лт}\,k}$. Therefore, in light of the statistical independence of the errors of individual channels, the probability that, given signal $u_1(t)$ presence, the threshold will not be exceeded in any one of the channels equals $(1 - P_{\text{лт}\,k})^{m-1}$; the probability that it, in addition, will not be exceeded in the first channel equals $P_{\text{пр}\,k}(1 - P_{\text{лт}\,k})^{m-1}$; consequently,

$$P_{\text{пр}} = P_{\text{пр}\,k}(1 - P_{\text{лт}\,k})^{m-1}. \tag{11.15}$$

We now will find distortion probability $P'_{\text{иск}}$. In this case, it is more convenient first to find the probability of no distortion

$$P'_{\text{с прав}} = 1 - P'_{\text{иск}}, \tag{11.16}$$

where $P'_{\text{с прав}}$ and $P'_{\text{иск}}$ -- conditional probabilities determined for the condition that one non-zero signal (it is immaterial exactly which one) is present at system input.

Consequently, $P'_{\text{с прав}}$ is the probability of a correct response for the condition that only one of the non-zero signals is present at system input. Evidently,

$$P'_{\text{с прав}} = P_{\text{c}}(u_1)\,P_{1\,\text{прав}} + P_{\text{c}}(u_2)\,P_{2\,\text{прав}} + \ldots + P_{\text{c}}(u_m)\,P_{m\,\text{прав}}, \tag{11.17}$$

where $P_{k\,\text{прав}}$-- probability of a correct response for the condition that signal u_k is present at system input.

In light of the aforementioned assumptions concerning the signal nature, this condition must be met

$$P_{1\,\text{прав}} = P_{2\,\text{прав}} = \ldots = P_{m\,\text{прав}}.$$

Considering this relationship and normality condition (11.13), we obtain:

$$P'_{\text{с прав}} = P_{1\,\text{прав}}, \qquad (11.18)$$

i. e., it is possible to compute correct response probability $P'_{\text{с прав}}$ in this case for the condition that signal $u_1(t)$ is present at system input.

In order for a correct response to be supplied here, it is necessary that threshold U_0 be exceeded in the system's first channel (Figure 11.6), while no threshold overrun occurs in all remaining channels. But, given presence of signal $u_1(t)$, the probability of threshold U_0 overrun in the first channel equals $(1 - P_{\text{пр}\,k})$, while the probability of a threshold non-overrun in all remaining channels equals $(1 - P_{\text{лт}\,k})^{m-1}$; therefore, the probability of coincidence /168
of all aforementioned events (considering their statistical independence) equals

$$P'_{\text{с прав}} = (1 - P_{\text{пр}\,k})(1 - P_{\text{лт}\,k})^{m-1}$$

and, considering (11.16), finally we obtain

$$P'_{\text{иск}} = 1 - (1 - P_{\text{пр}\,k})(1 - P_{\text{лт}\,k})^{m-1}. \qquad (11.19)$$

Formulas (11.11), (11.15), and (11.19) establish the link between the error probabilities in the system as a whole ($P_{\text{лт}}$, $P_{\text{пр}}$ and $P'_{\text{иск}}$) and in each channel of this system ($P_{\text{лт}\,k}$ and $P_{\text{пр}.k}$). Therefore, based on these formulas, it is possible with respect to the given (permissible) error probabilities in the system as a whole to determine permissible probabilities $P_{\text{лт}\,k}$ and $P_{\text{пр}\,k}$ in each channel, i. e., false-alarm and miss probabilities during simple binary detection. Based on magnitudes $P_{\text{лт}\,k}$ $P_{\text{пр}\,k}$ found, using the simple binary detection formulas [formula (5.31), for example], it is possible to determine requisite signal energy Q ($Q_{\text{ср}}$ for a fluctuating signal). Given such energy, error probabilities $P_{\text{лт}}$, $P_{\text{пр}}$ and $P'_{\text{иск}}$ occur in the quasi-optimum system (Figure 11.6).

As indicated above, given such signal energy, false-alarm and miss probabilities in the optimum system (Figure 11.1 or 11.5) will be exactly the same, while distortion probability $P_{\text{иск}}$ will satisfy inequality (11.8), i. e.,

$$P_{\text{неск}} \leqslant P'_{\text{неск}}. \tag{11.20}$$

Consequently, the computational procedure indicated above makes it possible relatively easily to determine probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$ in the optimum system and to find the upper limit for distortion probability $P_{\text{неск}}$. Especially-simple and clear relationships result if error probabilities $P_{\text{лт}}$, $P_{\text{пр}}$ and $P'_{\text{неск}}$ (and, consequently, $P_{\text{неск}}$) are sufficiently slight. Therefore, we will turn to examination of this case.

11.3 The Case of Slight Error Probabilities

We will assume that the false-alarm probability is sufficiently slight, and precisely is

$$P_{\text{лт}} \leqslant 0,1; \tag{11.21}$$

here, from (11.11) with less than 5% error we obtain:

$$P_{\text{лт}} = mP_{\text{лт}\,k} \text{ and } P_{\text{лт}\,k} = \frac{P_{\text{лт}}}{m}. \tag{11.22}$$

Formula (11.15) may be written in the form

$$P_{\text{пр}} = P_{\text{пр}\,k}\frac{1-P_{\text{лт}}}{1-P_{\text{лт}\,k}}.$$

Considering relationships (11.21) and (11.22), it is possible with less /169
than 10% error to assume

$$P_{\text{пр}} \approx P_{\text{пр}\,k}. \tag{11.23}$$

With about the same accuracy, from (11.19) we have

$$P'_{\text{неск}} \approx P_{\text{пр}} + P_{\text{лт}}\left(1-\frac{1}{m}\right). \tag{11.24}$$

Where m $\geqslant$ 10, it is possible to assume

$$P'_{\text{иск}} \approx P_{\text{пр}} + P_{\text{л т}}. \qquad (11.24a)$$

It follows from formulas (11.24) and (11.24a) that, for given probabilities $P_{\text{пр}}$ and $P_{\text{л т}}$, permissible probability $P'_{\text{иск}}$ already may not be selected randomly, but must be determined from these formulas. If you assume that $P_{\text{л т}} \leqslant 0{,}1$ and $P_{\text{пр}} \leqslant 0{,}1$, then here, the following always will be the case .

$$P'_{\text{иск}} \leqslant 0{,}2. \qquad (11.25)$$

As indicated above, finding requisite signal energy means finding permissible probabilities $P_{\text{пр}\,k}$ and $P_{\text{л т}\,k}$ in a simple binary detection channel using givens $P_{\text{пр}}$ and $P_{\text{л т}}$, then substituting them into the corresponding simple binary detection formulas.

In this case, as follows from (11.22) and (11.23), we obtain

$$P_{\text{пр}\,k} = P_{\text{пр}}; \quad P_{\text{л т}\,k} = \frac{P_{\text{л т}}}{m}; \qquad (11.26)$$

thus, the formula for signal energy for detection with recognition of m possible signals may be obtained from the corresponding simple binary detection formulas, if $P_{\text{л т}}$ is replaced by $P_{\text{л т}}/m$ in the latter.

The following simple binary detection formulas were obtained in preceding chapters:

a) for a precisely-known signal,

$$\frac{Q}{N_0} = \left(\sqrt{\ln \frac{1}{P_{\text{л т}}} - 1{,}4} + \sqrt{\ln \frac{1}{P_{\text{пр}}} - 1{,}4}\right)^2; \qquad (5.31a)$$

b) for a random-phase signal

$$\frac{Q}{N_0} = \left(\sqrt{\ln \frac{1}{P_{\text{л т}}}} + \sqrt{\ln \frac{1}{P_{\text{пр}}} - 1{,}4}\right)^2; \qquad (9.49)$$

c) for a fluctuating signal

$$\frac{Q_{\text{ср}}}{N_0} = \frac{1}{P_{\text{пр}}} \ln \frac{1}{P_{\text{лт}}}.$$

Replacing $P_{\text{лт}}$ with $P_{\text{лт}}/m$ in these formulas, we obtain the corresponding formulas for detection with recognition of m possible signals:

a) for precisely-known signals /170

$$\frac{Q}{N_0} = \left(\sqrt{\ln m + \ln \frac{1}{P_{\text{лт}}} - 1{,}4} + \sqrt{\ln \frac{1}{P_{\text{пр}}} - 1{,}4}\right)^2; \quad (11.27)$$

b) for random-phase (equally probable) signals

$$\frac{Q}{N_0} = \left(\sqrt{\ln m + \ln \frac{1}{P_{\text{лт}}}} + \sqrt{\ln \frac{1}{P_{\text{пр}}} - 1{,}4}\right)^2; \quad (11.28)$$

c) for fluctuating signals

$$\frac{Q_{\text{ср}}}{N_0} = \frac{1}{P_{\text{пр}}} \left(\ln m + \ln \frac{1}{P_{\text{лт}}}\right). \quad (11.29)$$

These formulas make it possible to determine energy Q (or $Q_{\text{ср}}$) of each signal $u_k(t)$ required for detection with recognition of m possible signals with given error probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$.

Since values $P_{\text{лт}}$ and $P_{\text{пр}}$ are identical for optimum and quasi-optimum systems, then formulas (11.27)--(11.29) in equal measure are valid for both systems.

The difference between optimum and quasi-optimum systems, as was demonstrated in § 11.2, consists only of the fact that, for given magnitudes $P_{\text{лт}}$ and $P_{\text{пр}}$, the distortion probability in a quasi-optimum system is determined from

formula (11.24), while distortion probability $P_{иск}$ satisfies inequalities (11.6) and (11.20), i. e., range from

$$P_{пр} \leqslant P_{иск} \leqslant P_{пр} + \left(1 - \frac{1}{m}\right) P_{л\,т}. \tag{11.30}$$

From formulas (11.24) and (11.30) we have

$$1 \leqslant \frac{P'_{иск}}{P_{иск}} \leqslant 1 + \left(1 - \frac{1}{m}\right) \frac{P_{л\,т}}{P_{пр}}. \tag{11.31}$$

It follows from these relationships that the difference between the distortion probabilities in quasi-optimum and optimum systems is less, the less ratio $P_{л\,т}/P_{пр}$, and, given $P_{л\,т}/P_{пр} \leqslant 0{,}1$, it is possible to assume:

$$P_{иск} \approx P'_{иск}. \tag{11.32}$$

This result is very understandable. Actually, it was demonstrated in § 11.2 that resultant $P'_{иск}$ is greater than $P_{иск}$ only because, in a quasi-optimum system, obtaining the correct response when signal $u_k(t)$ is present requires, along with everything else, that output voltages of all receivers $П_1, \ldots, П_m$, with the exception of $П_k$, not exceed threshold U_0 (Figure 11.6). In the optimum system (Figure 11.1), there is no requirement that this condition be met in order to obtain the correct response. Hence, it follows that the shortcoming of a quasi-optimum system compared with an optimum system (i. e., the $P'_{иск}$ overrun over $P_{иск}$) must be greater, the lower threshold U_0. Where $U_0 \longrightarrow \infty$, this shortcoming asymptotically /171 disappears, but, when $U_0 = 0$, the quasi-optimum system will become completely useless. Actually, if $U_0 = 0$ is selected, then a U_0 threshold overrun essentially always occurs in all Figure 11.6 system channels (i. e., dips in all indicators $I_1, \ldots, I_m$) and, as a result, it almost always is impossible to indicate which possible signal u_k is present at input.

Thus, the lower U_0, the greater ratio $P'_{иск}/P_{иск}$ must be. But, the lower U_0, the greater ratio $P_{л\,т}/P_{пр}$. Actually, if $U_0 = 0$, then false alarms always occur in the system and there are no misses, i. e., $P_{л\,т}/P_{пр} \to \infty$. If $U_0 \longrightarrow \infty$, then

there always are misses in the system and there are no false alarms, i. e., $P_{\text{лт}}/P_{\text{пр}} \to 0$. Consequently, when bias U_0 changes from infinity to zero, ratio $P_{\text{лт}}/P_{\text{пр}}$ rises from zero to infinity and, with U_0 inspection, it is possible to obtain any ratio $P_{\text{лт}}/P_{\text{пр}}$ value.

Thus, it follows from a purely qualitative examination that an increase in ratio $P_{\text{лт}}/P_{\text{пр}}$ means that ratio $P'_{\text{иск}}/P_{\text{иск}}$ also must increase. Inequality (11.31) provides quantitative confirmation of this result.

The ratios presented in this section make it possible very easily to determine requisite signal energy for both quasi-optimum and optimum systems.

Actually, knowing the number m of possible signal values and having been given permissible error probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$, it is possible to use a simple binary detection formula for the appropriate signal type [one of the (11.27)--(11.29) formulas, for instance] to determine requisite signal energy Q (or $Q_{\text{ср}}$). Given such energy, distortion probability is determined in a quasi-optimum system from formula (11.24) and from inequality (11.30) for an optimum system. Inequality (11.30) in the general case does not permit determination of a precise distortion probability value in an optimum system since it indicates only the limits in which the value is bounded. However, in many practical instances, knowledge of these boundaries turns out to be fully sufficient. Thus, for example, it follows from formula (11.30) that, where $P_{\text{лт}} \leqslant 0.1 P_{\text{пр}}$, it is possible with less than 10% error to assume:

$$P_{\text{иск}} = P'_{\text{иск}} = P_{\text{пр}} + \left(1 - \frac{1}{m}\right) P_{\text{лт}}. \tag{11.32a}$$

There is absolutely no requirement for precise knowledge of magnitude $P_{\text{иск}}$ for several signal types and, when determining requisite signal energy, we will assume a variation in magnitude $P_{\text{иск}}$ selection by a factor of several unities and more. In particular, this postulation occurs in the case of precisely-known and random-phase signals. Actually, formulas (11.27) and (11.28) are valid for these signal types and it is possible to assume (where $P_{\text{лт}} \leqslant 0{,}1$ and $P_{\text{пр}} \leqslant 0{,}1$) with slight error:

$$\frac{Q}{N_0} \approx \left(\sqrt{\ln m + \ln \frac{1}{P_{лт}}} + \sqrt{\ln \frac{1}{P_{пр}}} \right)^2. \tag{11.33}$$

We will assume /172

$$\frac{P_{лт}}{P_{пр}} = \gamma, \tag{11.34}$$

where γ -- some random number. Then, it follows from (11.30) that

$$P_{пр} \leqslant P_{иск} \leqslant \left[1 + \gamma \left(1 - \frac{1}{m} \right) \right] P_{пр}. \tag{11.35}$$

Using inequality (11.35) for magnitude $P_{иск}$ determination, let us make the maximum possible error, i. e., instead of assuming

$$P_{иск} = P_{пр}, \tag{11.36a}$$

we assume

$$P_{иск} = \left[1 + \gamma \left(1 - \frac{1}{m} \right) \right] P_{пр}. \tag{11.36b}$$

We will clarify how this will be reflected in requisite signal energy. In the first case, from (11.33), (11.34), and (11.36a) we obtain:

$$\frac{Q}{N_0} = \left(\sqrt{\ln m + \ln \frac{1}{\gamma P_{иск}}} + \sqrt{\ln \frac{1}{P_{иск}}} \right)^2 = \frac{Q_1}{N_0}.$$

In the second case, from formulas (11.33), (11.34), and (11.36b) we have

$$\frac{Q}{N_0} = \left\{ \sqrt{\ln m + \ln \frac{1}{\gamma P_{иск}} + \ln \left[1 + \gamma \left(1 - \frac{1}{m} \right) \right]} + \right.$$
$$\left. + \sqrt{\ln \frac{1}{P_{иск}} + \ln \left[1 + \gamma \left(1 - \frac{1}{m} \right) \right]} \right\}^2 = \frac{Q_2}{N_0}.$$

Consequently,

$$\frac{Q_2}{Q_1}=\left\{\frac{\sqrt{\ln m+\ln\frac{1}{\gamma P_{\text{иск}}}+\ln\left[1+\gamma\left(1-\frac{1}{m}\right)\right]}+\sqrt{\ln\frac{1}{P_{\text{иск}}}+\ln\left[1+\gamma\left(1-\frac{1}{m}\right)\right]}}{\sqrt{\ln m+\ln\frac{1}{\gamma P_{\text{иск}}}}+\sqrt{\ln\frac{1}{P_{\text{иск}}}}}\right\}^2. \tag{11.37}$$

It was assumed during derivation of this formula that, in accordance with (11.34)

$$\gamma P_{\text{нр}}=P_{\text{лт}}.$$

Since $P_{\text{лт}}\leqslant 1$ is always the case, then $\gamma P_{\text{нр}}\leqslant 1$ must be true. Considering formula (11.36a), we have

$$\gamma\leqslant\frac{1}{P_{\text{иск}}}. \tag{11.38}$$

Consequently, formula (11.37) is valid only for those parameter γ /173 values that satisfy condition (11.38).

Formula (11.37) demonstrates the factor by which requisite signal energy changes if the maximum possible error is made when determining distortion probability, i. e., the upper rather than lower limit is selected in inequality (11.35).

It follows from formula (11.37) that, the smaller permissible distortion probability $P_{\text{иск}}$, the less the difference between Q_2 and Q_1, i. e., the less impact the error has on magnitude $P_{\text{иск}}$ determination. The Figure 11.7 curves are based on formula (11.37). It is evident from the curves that, where $P_{\text{иск}}$ = 0.1 and m = 2, the error in requisite energy determination does not exceed 1 dB

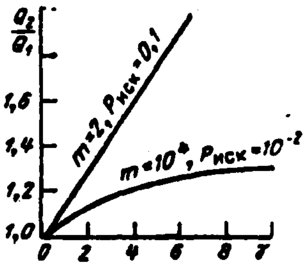


Figure 11.7

$\left(\frac{Q_2}{Q_1} \leqslant 1{,}26\right)$, if $\gamma \leqslant 1.8$. In a case more typical for radar ($m = 10^4$, $P_{иск} = 10^{-2}$), the error does not exceed 1 dB where $\gamma \leqslant 6$. It is evident from these examples that it is possible to use relationship (11.32a) for precisely-known and for random-phase signals, i. e., to assume

$$P_{иск} \approx P_{пр} + P_{лт}\left(1 - \frac{1}{m}\right) \qquad (11.39)$$

not only where $\frac{P_{лт}}{P_{пр}} \ll 1$, but also where $\frac{P_{лт}}{P_{пр}} \leqslant 1{,}8$ or even where $\frac{P_{лт}}{P_{пр}} \leqslant 6$.

Thus, for many cases interesting from a practical point of view, the energy required in a quasi-optimum system to provide given error probabilities is approximately the same as that in an optimum system and, consequently, it is possible to use the same computational formulas for both systems.

As already noted in § 11.2, the error probablities in the case of an optimum m-channel system (Figure 11.5) are the same as in the case of an optimum system of detection with recognition of m-orthogonal signals (Figure 11.1). Therefore, computational formulations presented above, formulas (11.27)--(11.32) in particular, are valid for an optimum m-channel system as well (Figure 11.5).

In conclusion, it should be underscored that it was assumed throughout this section that, in the case of m possible signals $u_1(t), \ldots, u_m(t)$ at system input (Figure 11.1), only one of these signals may be present during each sequence.

This is equivalent in the case of an m-channel system (Figure 11.5) to the fact that a signal may be present only in one of its channels.

In a more complex case where signals may be present simultaneously in several channels, results may differ significantly. For instance, in an m-channel system (Figure 11.5) let signals appear independently in any number of channels of this system. This signifies that, if it becomes known, for example, that there /174 is a signal at input of the k-th channel, then this will not impact in any way on the probability of signal presence or absence in the remaining channels. Since noise in different channels also is assumed to be statistically independent, then information on the nature of the voltages in the remaining channels plays no role whatsoever in solving the problem of whether or not a signal is present in a given channel.

Consequently, optimum solution of the overall problem requires each channel to provide an independently-optimum solution of the simple binary detection problem, i. e., to operate in the same mode, as occurs in the Figure 11.6 system. This denotes that, for the given conditions, the Figure 11.6 system is converted from a quasi-optimum into an optimum system (the assumption here is that there are signals in those channels where threshold U_0 is exceeded, i. e., indicators I_k dip, and there are no signals in the remaining channels). The Figure 11.1 system is based on selection of the greatest U_k voltage and, for the given conditions, is unacceptable.

11.4 Signal Detection With Recognition Compared To Complex Binary Detection

We will compare signal energy required for detection with recognition of m possible non-zero signals to the energy required in complex binary detection of such signals.

Since, in the second case, the requirement is only signal detection, without recognition of exactly which of the signals is present (if a signal is present), then a smaller volume of information will be subject to reproduction (reception) than in the first case. Therefore, solution of such a problem, all other conditions being equal, requires less signal energy than for simultaneous signal detection with recognition. We will clarify what savings in signal energy may be obtained

in the transition from signal detection with recognition to detection without recognition. Here, we will assume that all possible signals are mutually orthogonal and permissible error probabilities are not too great, but precisely are:

$$P_{лт} \leqslant 0{,}1,\ P_{пр} \leqslant 0{,}1 \text{ and } P_{нсн} \leqslant 0{,}1.$$

Then, the following formulas from Chapter 10 are valid for complex binary detection:

a) for precisely-known signals

$$\frac{Q'}{N_0} = \left(\sqrt{\ln\frac{1}{P_{лт}} - 1{,}4} + \sqrt{\ln\frac{1}{P_{пр}} - 1{,}4}\right)^2 + 0{,}5\ln m; \tag{10.16}$$

b) for random-phase signals

$$\frac{Q'}{N_0} = \left(\sqrt{\ln\frac{1}{P_{лт}}} + \sqrt{\ln\frac{1}{P_{пр}} - 1{,}4}\right)^2 + 0{,}5\ln m. \tag{10.18}$$

The following result during simultaneous signal detection and recognition: /175

a) for precisely-known signals

$$\frac{Q}{N_0} = \left(\sqrt{\ln m + \ln\frac{1}{P_{лт}} - 1{,}4} + \sqrt{\ln\frac{1}{P_{пр}} - 1{,}4}\right)^2; \tag{11.27}$$

b) for random-phase signals

$$\frac{Q}{N_0} = \left(\sqrt{\ln m + \ln\frac{1}{P_{лт}}} + \sqrt{\ln\frac{1}{P_{пр}} - 1{,}4}\right)^2. \tag{11.28}$$

For both signal types

$$P_{пр} \leqslant P_{нсн} \leqslant P_{пр} + \left(1 - \frac{1}{m}\right) P_{лт}. \tag{11.30}$$

We will assume for comparison of detection to simultaneous detection and recognition that, in both cases, identical detection error probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$ are permissible. Here, signal recognition distortion probability $P_{\text{иск}}$ is determined from relationship (11.30).

Then, from formulas (10.16), (10.18), (11.27), and (11.28), we will obtain

a) for precisely-known signals

$$\frac{Q}{Q'}=\frac{\left(\sqrt{\ln m+\ln\frac{1}{P_{\text{лт}}}-1,4}+\sqrt{\ln\frac{1}{P_{\text{пр}}}-1,4}\right)^2}{\left(\sqrt{\ln\frac{1}{P_{\text{лт}}}-1,4}+\sqrt{\ln\frac{1}{P_{\text{пр}}}-1,4}\right)^2+0,5\ln m}; \tag{11.40}$$

b) for random-phase signals

$$\frac{Q}{Q'}=\frac{\left(\sqrt{\ln m+\ln\frac{1}{P_{\text{лт}}}}+\sqrt{\ln\frac{1}{P_{\text{пр}}}-1,4}\right)^2}{\left(\sqrt{\ln\frac{1}{P_{\text{лт}}}}+\sqrt{\ln\frac{1}{P_{\text{пр}}}-1,4}\right)^2+0,5\ln m} \tag{11.41}$$

Ratio Q/Q' demonstrates the factor by which requisite signal energy increases during the transition from detection to simultaneous detection and recognition.

To simplify the analysis, initially we will accept that the follwing relationships (which usually are the case in practice) are satisfied:

$$\ln\frac{1}{P_{\text{лт}}}\gg 1,4 \quad\text{and}\quad \ln\frac{1}{P_{\text{лт}}}+\ln m\gg\ln\frac{1}{P_{\text{пр}}}. \tag{11.42}$$

Then, relationships (11.40) and (11.41) coincide and take the following form:

$$\frac{Q}{Q'}=1+\frac{1}{1+\frac{2\ln(1/P_{\text{лт}})}{\ln m}}. \tag{11.43}$$

Consequently, the result for any ratio $\frac{\ln(1/P_{лт})}{\ln m}$ is /176

$$1 \leqslant \frac{Q}{Q'} \leqslant 2,$$

i. e., discontinuance of recognition provides a signal energy savings of less than 3 dB.

If, instead of (11.42), one assumes

$$\ln\frac{1}{P_{лт}} = \ln\frac{1}{P_{пр}} \gg 1.4, \tag{11.44}$$

then, relationships (11.40) and (11.41) take the form

$$\frac{Q}{Q'} = \frac{4+2x+4\sqrt{1+x}}{8+x}, \tag{11.45}$$

where

$$x = \frac{\ln m}{\ln(1/P_{лт})}.$$

Here, the result again is

$$1 \leqslant \frac{Q}{Q'} \leqslant 2.$$

Finally, if one assumes in addition to condition (11.42) [or to (11.44)] that

$$\ln\frac{1}{P_{лт}} \gg \ln m, \tag{11.46}$$

then, from (11.43) [or from (11.45)], we will obtain $Q/Q' \approx 1$.

Thus, the following basic results are obtained for precisely-known and random-phase signals:

1. In the majority of cases of practical interest [i. e., when condition (11.42) or (11.44) is met], discontinuance of signal recognition provides a savings in requisite signal energy less by a factor of 2.

2. The result for a sufficiently-slight false-alarm probability [and, precisely when condition (11.46) is met] is $Q \approx Q'$, i. e., discontinuance of signal recognition does not provide any savings in requisite signal energy.

The second result, as L. R. Dobrushin pointed out [79], occurs for fluctuating signals as well, i. e., in the case of fluctuating signals, it also is possible to assume that $Q \approx Q'$ if false-alarm probability $P_{лт}$ is sufficiently slight.

SPECIAL FEATURES OF PULSE TRAIN ("PACKET") DETECTION AND RECOGNITION

12.1 General Nature of Pulse Trains

Let possible signal $u_c(t)$ be a train of n pulses, as depicted in Figure 12.1. For brevity, this signal usually is referred to as a pulse packet.

Each pulse $u_i(t)$ of such a packet completely is characterized by amplitude a_i, frequency f_i, moment of onset t_i, initial phase at moment of onset φ_i, and

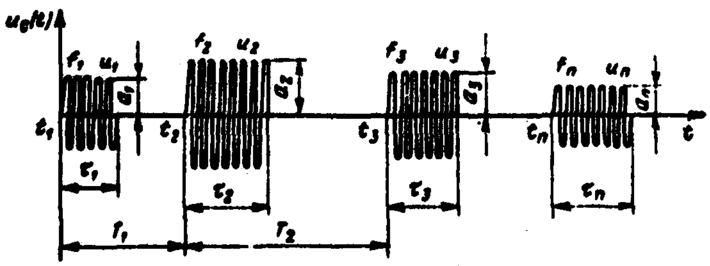


Figure 12.1

duration τ_i. If the relationships between the parameters of all packet pulses

are precisely known at the point of reception, then such pulses and such a pulse packet are called coherent.* Otherwise, the packet is called incoherent. It follows from determination of a coherent packet that, if even one of the packet pulses becomes known at the point of reception, for example parameters a_1, f_1, φ_1, and τ_1 of the first pulse, then entire packet $u_c(t)$ thereby will become known.

Consequently, a coherent pulse packet has exactly the same maximum possible number of unknown parameters as does the signal, which comprises one pulse. This circumstance makes it possible in a number of cases very simply to find the formula for the packet signal from the corresponding formulas for a monopulse signal.

The maximum possible number of unknown parameters in an incoherent pulse packet (of the type depicted in Figure 12.1) is very large and equals 5n.

The special features arising during binary detection (simple and complex) and during detection with recognition of various packet types are examined in this chapter.

12.2 Precisely-Known Packet Detection and Recognition /178

A packet is precisely known if the type of packet signal $u_c(t)$ is precisely known. The only unknown subject to determination in this case is whether or not a packet is present. Evidently, in order for packet $u_c(t)$ to be precisely known, it must be not only coherent, but all parameters of one of this packet's pulses, such as a_1, f_1, φ_1, and τ_1, also must be precisely known.

A precisely-known packet is a particular precisely-known signal and all the formulas presented above for detection (simple and complex), recognition, and simultaneous detection and recognition of precisely-known signals presented above are valid for it since, during its derivation, signal $u_c(t)$ was written in a general form, i. e., without concretization of its shape.

*From the word stsepleniye--coherence.

In particular, all Chapter 5 formulas and formulas (10.12) and (11.27) apply here. Signal $u_c(t)$ energy Q included in them equals

$$Q=\int_0^T u_c^2(t)\,dt; \tag{12.1}$$

therefore, as follows from Figure 12.1,

$$Q=\sum_{i=1}^{n} Q_i, \tag{12.1a}$$

where Q_i -- energy of the i-th pulse.

Consequently, all the aforementioned error probability formulas remain valid for precisely-known packets if Q in them is understood to mean the energy of the entire pulse packet equal to the sum of the energies of all n packet pulses [formula (12.1a)]. Hence, it follows that, when a precisely-known analog signal is split into a series of precisely-known pulses, signal detection and recognition errors in all the specific cases examined above remain unchanged if, during the split, total signal energy remains unchanged. In other words, for simple binary detection, and in a case of orthogonal signals also for complex binary detection, recognition (without detection), and simultaneous detection and recognition, the total requisite energy Q of each signal $u_k(t)$ does not change when an analog (monopulse) signal is split into a series of precisely-known pulses. However, such splitting complicates signal shape, with the result that the structure of the optimum receiver (correlator or optimum linear filter structure) is complicated accordingly. Therefore, if a system designer has the capability to select signal shape, then splitting it in the aforementioned cases into a series of pulses must be validated by some other significant advantages. In radar, for example, one advantage of splitting a monopulse sinusoidal signal into a series of pulses is the increase in system range resolution due to the reduction in the duration of each elementary /179 pulse. In addition, the conclusion made above concerning requisite signal energy invariability will apply to noise in the form of normal white noise and may not be valid for other types of noise.

12.3 Inverse Probability of a Random-Amplitude Random-Phase Packet

Let packet $u_c(t)$ be precisely known, with the exception of amplitude a_i and phase φ_i*. Parameters a_i and φ_i during each sequence [observation cycle (0, T)] are considered constant and, from one sequence to another, change like analog random magnitudes with n-dimensional a priori distributions $P(a_1, \ldots, a_n)$ and $P(\varphi_1, \ldots, \varphi_n)$, respectively. Conseouently, in this case, signal $u_c(t)$ has 2n random parameters and, in accordance with general formulas (8.11)--(8.14), its inverse probability may be written in the form

$$P_y(a_1, \ldots, a_n;\ \varphi_1, \ldots, \varphi_n) = \\ = k_1 P(a_1, \ldots, a_n)\, P(\varphi_1, \ldots \varphi_n)\, W_{\text{ш}}[y(t) - u_c(t)], \tag{12.2}$$

where

$$u_c(t) = u_c(a_1, \ldots, a_n;\ \varphi_1, \ldots, \varphi_n;\ t).$$

It is accepted here that there is no statistical dependence between amplitude and phase changes.

In addition, we will accept that the limits of change of all phases φ_i are identical and equal $0 - 2\pi$. Then, based on (8.11), it is possible to write

$$P_y(a_1, \ldots, a_n) = k_1 P(a_1, \ldots, a_n) \int_0^{2\pi} \cdots \int_0^{2\pi} P(\varphi_1, \ldots, \varphi_n) \times \\ \times W_{\text{ш}}[y(t) - u_c(t)]\, d\varphi_1 \ldots d\varphi_n. \tag{12.3}$$

---

*If moments t_i of pulse onset are not known precisely, but with errors Δt_i, slight in comparison with pulse durations τ_i, and if a large number of sinusoid periods ($f_i \tau_i \gg 1$) will be contained in each pulse, then it is possible to assume that moments t_i are precisely known and only phases φ_i are unknown and random.

In other words, in this case, the ambiguity of pulse onset moments is considered sufficiently completely by the fact that initial phases φ_i are considered ambiguous.

For noise in the form of normal white noise, in accordance with (1.25), we get

$$P_y(a_1, \ldots, a_n) = k_2 P(a_1, \ldots, a_n) \int_0^{2\pi} \cdots \int_0^{2\pi} P(\varphi_1, \ldots, \varphi_n) \times$$
$$\times \exp\left[-\frac{1}{N_0} \int_0^T [y(t) - u_c(t)]^2\, dt\right] d\varphi_1 \ldots d\varphi_n.$$

Expanding the square brackets under the integral, we have /180

$$P_y(a_1, \ldots, a_n) = k_3 P(a_1, \ldots, a_n)\, e^{-Q/N_0} \int_0^{2\pi} \cdots \int_0^{2\pi} P(\varphi_1, \ldots, \varphi_n) \times$$
$$\times \exp\left[-\frac{2}{N_0} \int_0^T y(t)\, u_c(t)\, dt\right] d\varphi_1 \ldots d\varphi_n, \tag{12.4}$$

where Q is determined from formulas (12.1) and (12.1a), while k_3 -- normalizing coefficient.

Further, we will examine several more-particular cases.

First case. Packet pulses are coherent among themselves. This signifies that there is a composite statistical link between individual pulse parameters and the n-dimensional probability densities in formula (12.4) may be replaced by unidimensional distributions for any of the packet pulses, for the first pulse, for example. In addition, we will accept that frequencies f of all pulses are identical, while distribution of phases is uniform. Then, formula (12.4) may be written in the form

$$P_y(a_1) = k_4 P(a_1)\, e^{-Q/N_0} \int_0^{2\pi} e^{\eta(a_1, \varphi_1)}\, d\varphi_1, \tag{12.5}$$

where

$$\eta(a_1, \varphi_1) = \frac{2}{N_0} \sum_{l=1}^{n} \int_{t_l}^{t_l+\tau_l} y(t)(a_1 + \Delta a_l) \cos(\omega t + \varphi_1 + \Delta\varphi_l)\, d\varphi_1;$$
$$\Delta a_l = a_l - a_1; \quad \Delta\varphi_l = \varphi_l - \varphi_1.$$

Since the pulse packet is coherent, the Δa_i and $\Delta \varphi_i$ are precisely-known magnitudes.

<u>Second case</u>. Packet pulse amplitudes are linked among themselves by a precisely-known dependence, while phases φ_i are statistically independent and distributed uniformly (ranging from 0 to 2π).

In this case, it is possible in formula (12.4) to replace $P_y(a_1, \ldots, a_n)$ and $P(a_1, \ldots, a_n)$ by $P_y(a_1)$ and $P(a_1)$, respectively and also to assume

$$P(\varphi_1, \ldots, \varphi_n) = P(\varphi_1) \ldots P(\varphi_n) = \left(\frac{1}{2\pi}\right)^n;$$

then formula (12.4) takes the form

$$P_y(a_1) = k_n P(a_1) e^{-\frac{Q}{N_0}} \int_0^{2\pi} \ldots \int_0^{2\pi} e^{\frac{2}{N_0}\int_0^T y(t) u_c(t)\, dt}\, d\varphi_1 \ldots d\varphi_n. \qquad (12.6)$$

But

$$\int_0^T y(t) u_c(t)\, dt = \sum_{i=1}^{n} \int_{t_i}^{t_i+\tau_i} y(t) a_i \cos(\omega t + \varphi_i)\, dt;$$

therefore /181

$$P_y(a_1) = k_n P(a_1) e^{-Q/N_0} \prod_{i=1}^{n} \int_0^{2\pi} e^{\eta(a_i, \varphi_i)}\, d\varphi_i, \qquad (12.7)$$

where

$$\eta(a_i, \varphi_i) = \frac{2a_i}{N_0} \int_{t_i}^{t_i+\tau_i} y(t) \cos(\omega t + \varphi_i)\, dt. \qquad (12.8)$$

Comparing expressions (9.9) and (9.10) with (12.7) and (12.8), it is not difficult to become convinced that, in this case

$$P_y(a_1) = k_6 P(a_1) e^{-Q/N_0} \prod_{i=1}^{n} I_0\left(\frac{2a_i M_i}{N_0}\right), \tag{12.9}$$

where

$$Q = \sum_{i=1}^{n} \frac{a_i^2}{2} \tau_i; \; a_i = a_1 + \Delta a_i$$

(Δa_i and τ_i -- precisely-known magnitudes);

$$\left.\begin{aligned} &M_i = \sqrt{X_i^2 + Y_i^2}; \\ &X_i = \int_{t_i}^{t_i+\tau_i} y(t) \cos \omega t \, dt; \quad Y_i = \int_{t_i}^{t_i+\tau_i} y(t) \sin \omega t \, dt. \end{aligned}\right\} \tag{12.9a}$$

Evidently, magnitude M_i plays the same role as parameter M in the case of a monopulse system and may be computed using identical approaches, using an optimum filter, for example. Since, during computation of M_i for various number i, the only thing that changes are integration limits, which do not overlap among themselves, then all M_i magnitudes may be obtained with the aid of an optimum filter matched only with first pulse $u_1(t)$ by means of the corresponding time shifts.

Consequently, when using formula (12.9) for inverse probability computation in this case, it suffices to have an optimum filter matched with single pulse $u_1(t)$ and the appropriate delay circuits.

It usually is more convenient to compute $\ln P_y(a_1)$ rather than $P_y(a_1)$; here, instead of formula (12.9) (assuming $k_6 = 1$), we have

$$\ln P_y(a_1) = \ln P(a_1) - \frac{Q}{N_0} + \sum_{i=1}^{n} \ln I_0\left(\frac{2a_i M_i}{N_0}\right). \tag{12.10}$$

Third case. Individual packet pulse amplitudes and phases statistically are independent; phases are distributed uniformly (ranging from 0 up to 2π).

In this case, in formula (12.4) it is possible to assume

$$P(a_1, \ldots, a_n) = P(a_1) \ldots P(a_n).$$

Otherwise, this case does not differ from the preceding one. Thus, /182
instead of (12.9), it is possible immediately to write:

$$P_y(a_1, \ldots, a_n) = k \prod_{i=1}^{n} P(a_i) e^{-\frac{a_i^2 \tau_i}{2N_0}} I_0\left(\frac{2a_i M_i}{N_0}\right), \tag{12.11}$$

where M_i is determined from the same formulas as in the preceding case.

12.4 Detection and Recognition of Coherent Random Initial Phase Packets

In this case, the analysis may be based on expression (12.5) for inverse probability. However, the basic relationships for a signal of this type already were presented in preceding chapters.

Actually, since all packet amplitudes are coherent among themselves, possible signal $u_k(t)$ [signal $u_c(t)$ in Figure 12.1] will fall into the formula (9.50) class in which functions a(t) and $\psi(t)$ are precisely known and describe the packet pulse amplitude and intitial phase change when this pulse's number i changes. Therefore, all formules presented in preceding sections for a signal with known amplitude and random equiprobable phase are valid for error probabilities if energy Q in these formulas is understood to mean the energy of a packet of n pulses, determined from formula (12.1). Thus, for example, formula (11.33) is valid for detection with recognition of equiprobable orthogonal packet signals with equal energies, to wit

$$\frac{Q}{N_0} \approx \left(\sqrt{\ln m + \ln \frac{1}{P_{n\tau}}} + \sqrt{\ln \frac{1}{P_{np}}}\right)^2, \tag{11.33}$$

where Q = nQ' — packet energy.

The structure of optimum receiver Π_k included in the detection system is determined from formula (9.51). Parameter M_1 included in this formula may be computed, as noted above, using an optimum filter and amplitude detector or by the Figure 9.1 computational device. In the latter case, cos ω t and sin ω t everywhere should be replaced by a(t) cos [ωt + ψ(t)] and a(t) sin [ω t + ψ(t)], respectively.

12.5 Coherent Fluctuating Packet Detection and Recognition

A coherent packet is referred to as fluctuating if amplitude a_1 and phase ϕ_1 from one sequence to another change as analog random magnitudes with a priori distributions $W(a_1)$ and $P(\phi_1)$, respectively. (Since the packet is coherent, then fluctuations in first pulse amplitude and phase fully determine the /183
fluctuation of the amplitudes and phases of all subsequent packet pulses).

Here

$$P(a_1) = P(C)\, W(a_1), \tag{12.12}$$

where P(C) -- a priori distribution of packet presence. In this case, the analysis also may be performed based on expression (12.5).

However, given certain assumptions, results may be obtained directly from formulas presented in preceding chapters. Thus, we will assume that packet pulses have identical amplitudes and phases ($a_i = a_0$, $f_i = f_0$) and amplitude is distributed based on Rayleigh's law, while phase is equally probable [amplitude and phase are unchanged within each observation cycle 0, T)]. Then, error probabilities are determined from the formulas presented in the preceding chapter for a fluctuating signal. Here, energy Q_{cp} should be understood to mean the average energy of the entire packet, i. e.,

$$Q_{cp} = \sum_{i=1}^{n} Q_{cp\,i}, \tag{12.13}$$

where $Q_{cp\,i} = a_0^2 \tau_i / 2$ -- average energy of the i-th packet pulse.

If all packet pulses have identical duration τ, then

$$Q_{cp} = nQ_{cp\,1}, \tag{12.13a}$$

where $Q_{cp\,1}$ -- average energy of one packet pulse. Thus, for example, during fluctuating packet detection with recognition, formula (11.29) is valid if, in it, Q_{cp} is understood to mean average packet energy determined from formula (12.13).

12.6 Detection and Recognition of Incoherent Packets With Known Amplitudes and Independent Random Phases

If packet pulse amplitudes are interlinked by a precisely-known relationship and phases Φ_i are statistically independent and uniformly distributed (ranging from 0 to 2π), then the second case examined in § 12.3 occurs here and analysis may be performed based on formula (12.10) for packet inverse probability.

Initially, we will examine simple binary detection of such a packet. Here, it should be assumed in formula (12.10) that amplitude a_1 either equals precisely-known magnitude a_{01} (if a packet is present), or is identical with zero (if no packet is present). Therefore, a response to the question of whether or not a packet is present requires comparison among themselves of inverse probabilities (or their logarithms) $P_y(C)$ and $P_y(0)$ packet presence or absence, respectively, where, in accordance with formula (12.10)

$$\left.\begin{aligned} &\ln P_y(C) = \ln P_y(a_{01}) = \ln P(C) - \frac{Q}{N_0} + \sum_{i=1}^{n} \ln I_0\left(\frac{2a_i M_i}{N_0}\right); \\ &\ln P_y(0) = \ln P(0), \end{aligned}\right| \tag{12.14}$$

and $P(C) = P(a_{01})$ is the a priori probablity of packet presence.

It follows from (10.5) and (12.14) that the decision must be that a signal is present if the following condition is met /184

$$\sum_{i=1}^{n} \ln I_0\left(\frac{2a_i M_i}{N_0}\right) > U_0. \tag{12.15}$$

where

$$U_0 = \frac{Q}{N_0} + \ln\left[\frac{P(0)}{P(C)}\eta\right]; \tag{12.16}$$

where η -- weight factor considering the relative danger of false alarms and signal misses.

If errors of both types are equally dangerous, i. e., minimum composite error probablity must be insured, then it should be assumed

$$\eta = 1$$

If condition (12.15) is not met, then the decision is made that no packet is present.

As noted above, parameter M_i may be computed (given certain restrictions that usually occur) by an optimum linear filter matched with pulse $u_i(t)$ and by an amplitude detector.

Since pulses $u_i(t)$ in the general case differ not only with respect to amplitude and time of action, but also with respect to carrier frequency f_i, then computation

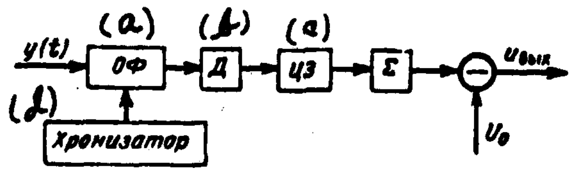


Figure 12.2. (a) -- OF [optimum filter]; (b) -- D [detector]; (c) -- TsZ [delay circuit]; (d) -- Timer.

of the sum included in inequality (12.15) requires n individual filters or one filter, but with parameters changing from pulse to pulse. In the latter case, the structural schematic of the optimum binary detection system has the form depicted in Figure 12.2. The timer insures optimum filter connection only for time intervals

corresponding to the time of action of the anticipated pulses, and also insures a change in filter amplification and tuning frequency from pulse to pulse in accordance with the changes in amplitudes a_i and frequencies of the individual packet pulses. If the anticipated pulses have identical amplitudes and frequencies, the requirement for the latter two operations is eliminated.

Amplitude detector *Д* with an ln $I_0(x)$ type response curve, delay circuit *ЦЗ* , adder Σ , and threshold bias U_0 are connected beyond the optimum filter. When threshold U_0 is exceeded, i. e., where $u_{sum} > 0$, the response is that a packet is present. Otherwise, the decision is that no packet is present.

For binary detection of a monopulse signal (n = 1), the response curve /185 shape, as noted in § 11.1, may be replaced by any monotonic curve, with appropriate threshold bias U_0 adjustment. In the given case (where $n \neq 1$), this postulation, strictly speaking, is invalid since the adding operation will occur prior to comparison with the threshold. However, even in this case, choice of the response curve shape is not very critical, especially if the number of pulses in the packet is small.

The probability of satisfying inequality (12.15) when there is and is not a signal present should be computed to find error probabilities. In the general case, solution of this problem encounters significant mathematical difficulties. Therefore, several assumptions have to be made.

First, we will assume that $n \gg 1$. Here, we may assume that the law of distribution of the sum at the left side of inequality (12.15) is normal since the terms making up this sum are statistically independent.

We will designate

$$y = \sum_{i=1}^{n} \ln I_0\left(\frac{2a_i M_i}{N_0}\right) = \sum_{i=1}^{n} z_i, \qquad (12.17)$$

where

$$z_i = \ln I_0\left(\frac{2a_i M_i}{N_0}\right); \qquad (12.18)$$

then, in light of the assumption made, it is possible to write the law of magnitude y distribution in the form

$$P(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-(y-\bar{y})^2/2\sigma_y^2}, \qquad (12.19)$$

where $\bar{y}$ and σ_y^2 -- magnitude y mean value and dispersion.

We will compute parameter $\bar{y}$ and σ_y^2 values when a signal is and is not present. It follows from (12.17) that

$$\bar{y} = \sum_{i=1}^{n} \bar{z}_i; \quad \sigma_y^2 = \sum_{i=1}^{n} \sigma_{z_i}^2, \qquad (12.20)$$

where $\bar{z}_i$ and $\sigma_{z_i}^2$ -- term z_i mean value and dispersion. Thus, it suffices for finding distribution P(y) to compute random magnitude z_i mean value and dispersion determined from formula (12.18).

We will designate

$$x_i = \frac{2a_i M_i}{N_0}; \qquad (12.21)$$

then

$$z_i = \ln I_0(x_i). \qquad (12.22)$$

We will assume that /186

$$x_i \ll 1; \qquad (12.22a)$$

then, considering the properties of function $I_0(x_i)$, we will obtain

$$z_i \approx \frac{x_i^2}{4} = \left(\frac{a_i M_i}{N_0}\right)^2. \tag{12.23}$$

Consequently,

$$\bar{z}_i = \overline{\left(\frac{a_i M_i}{N_0}\right)^2}.$$

Considering (9.13), the result is

$$\bar{z}_i = \frac{a_i^2}{N_0^2}(\overline{X_i^2} + \overline{Y_i^2}). \tag{12.24}$$

But, we found in § 9.2 that magnitudes $\frac{2a}{N_0}X$ and $\frac{2a}{N_0}Y$ when there is a signal are independent random magnitudes with zero mean values and with dispersions equalling $2Q/N_0$. Consequently, in the case under examination, this must be the case

$$\overline{\left(\frac{2a_i}{N_0}X_i\right)^2} = \overline{\left(\frac{2a_i}{N_0}Y_i\right)^2} = \frac{2Q_i}{N_0}.$$

Therefore, when there is no signal (i. e., only noise), formula (12.24) provides

$$\bar{z}_i = \frac{Q_i}{N_0}. \tag{12.25}$$

It is convenient to use the following approach to find dispersion $\sigma_{z_i}^2$ when a signal is and is not present, as well as to determine magnitude $\overline{z_i}$ when there is a signal.

Magnitude M_i, as was demonstrated above, is proportional to optimum linear

filter output voltage envelope $U_m(t)$ (at moment $t = T$). Therefore, it is possible to write expression (12.23) in the form

$$z_i = AU_m^2(T), \tag{12.26}$$

where A -- some constant.

This equation corresponds to the square-law detector response curve equation. Therefore, z_i may be considered square-law detector output voltage.

When there is no signal, only noise with a normal law of distribution arrives at the input of this detector. Here

$$\bar{z}_i = A\overline{U_m^2} = A2U_ш^2, \tag{12.27}$$

where $U_ш^2$ -- mean square of the noise voltage at detector input.

Dispersion $\sigma_{z_i}^2$ in this case is nothing other than the mean square of the noise voltage at square-law detector output, therefore

$$\sigma_{z_i}^2 = A^2 4U_ш^4. \tag{12.28}$$

When there is a signal, detector output voltage is the sum of normal /187
noise with extant value $U_ш$ and a harmonic signal with amplitude U_{mc}. Here, the result at square-law detector output is

$$\bar{z}_i = A2U_ш^2\left(1 + \frac{U_{mc}^2}{2U_ш^2}\right) \tag{12.29}$$

and

$$\sigma_{z_i}^2 = (A2U_ш^2)^2\left(1 + \frac{U_{mc}^2}{U_ш^2}\right). \tag{12.30}$$

It follows from comparison of relationships (12.25) and (12.27) that

$$A2U_{ш}^{2}=\frac{Q_l}{N_0}.\tag{12.31}$$

Ratio $U_{m\,c}/U_{ш}$ in formulas (12.29) and (12.30) is the ratio of signal amplitude (at moment t = T) to the mean square noise voltage at optimum linear filter output. Therefore, in accordance with formula (2.34), we have

$$\frac{U_{m\,c}^{2}}{U_{ш}^{2}}=\frac{2Q_l}{N_0}.\tag{12.32}$$

Substituting expressions (12.31) and (12.32) into formulas (12.27)--(12.30), we obtain:

a) when there is no signal

$$\bar{z}_l=\frac{Q_l}{N_0};\ \sigma_{z_l}^{2}=\frac{Q_l^{2}}{N_0^{2}};\tag{12.33}$$

b) when there is a signal

$$\bar{z}_l=\frac{Q_l}{N_0}\left(1+\frac{Q_l}{N_0}\right);\ \sigma_{z_l}^{2}=\frac{Q_l^{2}}{N_0^{2}}\left(1+\frac{2Q_l}{N_0}\right).\tag{12.34}$$

Further, we will assume for simplicity that all packet pulses have identical energies, i. e.,

$$Q_l=\frac{Q}{n},\tag{12.35}$$

where Q -- energy of the entire packet.

Then, fr. rmulas (12.20), (12.33), (12.34), and (12.35), we will obtain:

a) when there is no signal

$$\bar{y}=\frac{Q}{N_0};\quad \sigma_y^2=\frac{1}{n}\cdot\frac{Q^2}{N_0^2};\tag{12.36}$$

b) when there is a signal

$$\bar{y}=\frac{Q}{N_0}\left(1+\frac{Q}{nN_0}\right);\quad \sigma_y^2=\frac{Q^2}{nN_0^2}\left(1+\frac{2Q}{nN_0}\right).\tag{12.37}$$

Based on the resultant data, it is not difficult to compute false-alarm and signal miss probabilities.

Actually, it follows from relationships (12.15) and (12.17) that the /188 false-alarm probability is the probability that this inequality is satisfied

$$y>U_0$$

when there is no signal, i. e.,

$$P_{\pi\tau}=\int_{U_0}^{\infty}P(y)\,dy=\frac{1}{\sqrt{2\pi\sigma_y^2}}\int_{U_0}^{\infty}e^{-(y-\bar{y})^2/2\sigma_y^2}\,dy,\tag{12.38}$$

where $\bar{y}$ and σ_y^2 are determined by formulas (12.36).

Signal miss probability is the probability that this inequality is satisfied

$$y\leqslant U_0$$

when there is a signal, i. e.,

$$P_{\text{пр}}=\int_{-\infty}^{U_0}P(y)\,dy=\frac{1}{\sqrt{2\pi\sigma_y^2}}\int_{-\infty}^{U_0}e^{-(y-\bar{y})^2/2\sigma_y^2}\,dy,\tag{12.39}$$

where $\bar{y}$ and σ_y^2 are determined from formulas (12.37).

Through replacement of variables and considering formulas (12.36) and (12.37), it is easy to reduce expressions (12.38) and (12.39) to the following form:

$$P_{\text{лт}}=\frac{1}{\sqrt{2\pi}}\int_{\alpha_1}^{\infty} e^{-z^2/2}\,dz;\quad P_{\text{пр}}=\frac{1}{\sqrt{2\pi}}\int_{\alpha_2}^{\infty} e^{-z^2/2}\,dz, \tag{12.40}$$

where

$$\alpha_1=\sqrt{n}\,\frac{U_0-\frac{Q}{N_0}}{\frac{Q}{N_0}};\quad \alpha_2=\frac{\frac{Q}{N_0}\left(1+\frac{Q}{nN_0}\right)-U_0}{\frac{Q}{N_0\sqrt{n}}\sqrt{1+\frac{2Q}{nN_0}}}. \tag{12.41}$$

It may be demonstrated that it is advisible in formulas (12.41) to replace U_0 based on formula (12.16). However, this turns out to be inaccurate in this case.

Actually, formula (12.16) determines threshold U_0 optimum value, when supplied to it is normalized voltage of the type

$$u_{\text{вых}}=\sum_{i=1}^{n} z_i=\sum_{i=1}^{n}\ln I_0\left(\frac{2a_i M_i}{N_0}\right)$$

[see inequality (12.15)]. But, during derivation of formulas (12.40), we replaced this precise voltage $u_{\text{вых}}$ expression with an approximate expression of the type

$$u_{\text{вых}}=\sum_{i=1}^{n} z_i=\sum_{i=1}^{n}\left(\frac{a_i M_i}{N_0}\right)^2$$

and assumed that magnitude $u_{\text{вых}}$ has a normal law of distribution. Given such a changed (i. e., approximate rather than precise) form of the voltage /189
$u_{\text{вых}}$ distribution, the optimum threshold voltage U_0 value also turns out to

be somewhat different and must be determined, not from formula (12.16), but from the following formula:

$$U_0 = \frac{Q}{N_0} + \ln\left[\eta \frac{P(0)}{P(C)}\right] + \varepsilon. \qquad (12.16a)$$

As analysis shows, if $\frac{Q}{N_0} \gg \left|\ln\left[\frac{\eta P(0)}{P(C)}\right]\right|$, then correction ε has the following form:

$$\varepsilon \approx \frac{Q}{nN_0}\left(1 + \frac{Q}{2N_0}\right).$$

We will examine a case of slight error probabilities $P_{лт}$ and $P_{пр}$.

It follows from formulas (12.40) that probabilities $P_{лт}$ and $P_{пр}$ may be slight only if limits α_1 and α_2 are sufficiently large; but, given a sufficiently-large lower limit α, it is possible to use asymptotic expansion of the probability integral and to assume

$$\frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-z^2/2}\, dz \approx \frac{1}{\sqrt{2\pi}\,\alpha} e^{-\alpha^2/2};$$

therefore, instead of (12.40), it is possible to accept

$$P_{лт} \approx \frac{1}{\sqrt{2\pi}\,\alpha_1} e^{-\alpha_1^2/2}; \quad P_{пр} \approx \frac{1}{\sqrt{2\pi}\,\alpha_2} e^{-\alpha_2^2/2}, \qquad (12.42)$$

hence, we obtain:

$$\left.\begin{aligned} \alpha_1^2 &= 2\ln\frac{1}{P_{лт}} - 1{,}85 - 2\ln\alpha_1; \\ \alpha_2^2 &= 2\ln\frac{1}{P_{пр}} - 1{,}85 - 2\ln\alpha_2. \end{aligned}\right\} \qquad (12.43)$$

Since in this case α_1 and α_2 are sufficiently great, then it is possible in (12.43) in the first approximation to disregard terms 2 ln α_1 and 2 ln α_2 in comparison with α_1^2 and α_2^2, respectively. The result is:

$$\left.\begin{aligned}\alpha_1^2 &\approx 2\ln\frac{1}{P_{лт}} - 1{,}85;\\ \alpha_2^2 &\approx 2\ln\frac{1}{P_{пр}} - 1{,}85.\end{aligned}\right\} \qquad (12.44)$$

However, as the analysis shows, it is advisible for best approximation, instead of (12.44), to assume

$$\left.\begin{aligned}\alpha_1^2 &\approx 2\ln\frac{1}{P_{лт}} - 2{,}8;\\ \alpha_2^2 &\approx 2\ln\frac{1}{P_{пр}} - 2{,}8.\end{aligned}\right\} \qquad (12.45)$$

On the other hand, solving equation (12.41) for magnitude Q/N_0, we obtain /190

$$\frac{Q}{N_0} = (\sqrt{\alpha_1^2 n} + \alpha_2^2)\sqrt{(\sqrt{\alpha_1^2 n} + \alpha_2^2)^2 - (\alpha_1^2 - \alpha_2^2)\, n}. \qquad (12.46)$$

Formulas (12.45) and (12.46) make it possible quite simply to compute the packet energy Q required to insure permissible error probabilities $P_{лт}$ and $P_{пр}$ for a given number of pulses n.*

As computations show, where $P_{лт} \leqslant 0{,}1$ and $P_{пр} \leqslant 0{,}1$, the formula (12.46) error does not exceed 0.5 db, compared to more-precise integral relationships (12.40). However, it should be remembered that both formula (12.46) and relationships (12.40) are sufficiently precise only where $n \gg 1$.

*It follows from relationships (12.41), (12.45), and (12.46) that, for given error probabilities $P_{лт}$ and $P_{пр}$, threshold U_0 magnitude must be determined from requisite magnitude α_1, i. e., from formula $U_0 = \frac{Q}{N_0} + \frac{\alpha_1 Q}{N_0 \sqrt{n}}$, where Q/N_0 is determined from formula (12.46), while α_1 and α_2 are determined from formulas (12.45).

If n is very large, such as $n \gg \alpha_1^2$ and $n \gg \alpha_2^2$ or

$$n \gg (\alpha_1+\alpha_2)^2, \tag{12.47}$$

where α_1 and α_2 in turn are magnitudes greater than unity, then formula (12.46) takes the form

$$\frac{Q}{N_0} \approx (\alpha_1+\alpha_2)\sqrt{n}. \tag{12.48}$$

In addition, during derivation of formulas (12.46) and (12.48), the assumption was made [see (12.22a)] that $x_i \ll 1$. We will clarify the concept behind this assumption.

It follows from relationships (12.23), (12.33), and (12.34) that, when there is no signal

$$\overline{x_i^2} = 4\frac{Q_i}{N_0},$$

while, when there is a signal

$$\overline{x_i^2} = 4\frac{Q_i}{N_0}\left(1+\frac{Q_i}{N_0}\right);$$

consequently, instead of (12.22a), it is possible to write the following condition:

$$\frac{Q_i}{N_0} \ll 1, \tag{12.49}$$

i. e., the signal-to-noise power ratio for each (individual) packet pulse must

be sufficiently small. Here, the optimum response curve is a square-law curve (see equality (12.26)] and the conclusions made above are valid.

If condition (12.49) is not met, i. e., the signal-to-noise ratio for every packet pulse is not too low, then formulas (12.46) and (12.48) will remain valid if, as usual, one assumes that the response curve is a square-law curve. However, in this case, such a curve already will not be optimum since the optimum /191 characteristic is described by equation (12.22) and, given suffiently-large x_i, will strive towards straight line

$$z_i = x_i, \tag{12.50}$$

i. e., towards the characteristic of a linear detector.

Hence, it follows that, when condition (12.49) is not met, somewhat better results (lower Q/N_0) may be obtained in the optimum receiver than is the case from formula (12.46).

During formula (12.46) derivation, we assumed that the law of distribution of the (12.17) sum was normal, which is valid only for very large n values. Kaplan obtained more-precise results [16].* He also assumed a square-law response curve, i. e., in accordance with (12.17), he assumed that

$$y = \sum_{i=1}^{n} \frac{1}{4} x_i^2. \tag{12.51}$$

Here, when there is no signal, magnitude y has a chi-square distribution with 2n degrees of freedom.

Kaplan approximated this distribution as an expansion into a power series with respect to a normally-distributed magnitude, which, naturally, provides greater precision than the above P(y) distribution normalization.

*Useful relationships for a case of square-law detection [with respect to law (12.15)] will be found also in the V. S. Sragovich article [123].

Kaplan assumed distribution P(y) as normal for a case of signal-plus-noise activity since, in this case, on the one hand, distribution P(y) is closer to a normal distribution (than when there is no signal) and, on the other hand, it turns out to be more complicated to obtain a more-precise approximation here.

The formula Kaplan obtained, using our designations, may be reduced to the following form:

$$\frac{Q}{N_0}=\left(\frac{\alpha_1^2}{3}+\alpha_2^2\right)+ \\ +\sqrt{n}\left[\alpha_1+\alpha_2\sqrt{1+\frac{2\alpha_1}{\sqrt{n}}+\frac{\frac{2}{3}\alpha_1^2+\alpha_2^2}{n}}\right], \tag{12.52}$$

where α_1 and α_2 are determined by formulas (12.45) (where $P_{лт}\leqslant 0{,}1$ and $P_{пр}\leqslant 0{,}1$).

When n $\gg$ 1, formulas (12.52) and (12.46), as could be expected, coincide and boil down to very simple expression (12.48). However, if n is not too large, i. e., condition (12.47) is not met, more-precise formula (12.52) should be used.

However, this formula also only is valid when n values are not too small and its error increases with a decrease in n. In the worst case, when /192
n = 1, from formula (12.52) we obtain

$$\frac{Q}{N_0}=\frac{\alpha_1^2}{3}+\alpha_2^2+\alpha_1+\alpha_2\sqrt{1+2\alpha_1+\frac{2}{3}\alpha_1^2+\alpha_2^2}, \tag{12.53}$$

then, when in actuality n = 1, formula (9.49) is valid, which, considering (12.44), provides

$$\frac{Q}{N_0}\approx\frac{(\alpha_1+\alpha_2)^2}{2}. \tag{12.54}$$

It follows from comparing formulas (12.53) and (12.54) that, when these conditions are met

$$P_{np} \leqslant 0{,}1 \text{ and } P_{AT} \leqslant P_{np} \tag{12.55}$$

(i. e., when $\alpha_2 \geqslant 1.65$ and $\alpha_1 \geqslant \alpha_2$), which usually is the case, the formula (12.53) error, compared to formula (12.54), does not exceed 2 db. Hence, it follows that formula (12.52), precise where $n \gg 1$, also provides, for any slight changes in n, a relatively-small error not exceeding 2 dB.

It is possible to reduce this error further if formula (12.52) is adjusted in the following way.

Instead of (12.52), we will assume that

$$\frac{Q}{N_0} = \left(\frac{\alpha_1^2}{3} + \alpha_2^2\right) + \\ + \sqrt{n}\left[\alpha_1 + \alpha_2 \sqrt{1 + \frac{2\alpha_1}{\sqrt{n}} + \frac{\frac{3}{2}\alpha_1^2 + \alpha_2^2}{n}}\right] + A. \tag{12.56}$$

This expression differs from (12.52) only in additional term A, which does not depend on n; therefore, where $n \longrightarrow \infty$, formula (12.56) coincides with (12.52), i. e., as usual provides the correct result.

We will select magnitude A so that formula (12.56) will provide a correct result in another extreme case as well, i. e., when n = 1. Then, when n = 1, comparing the right sides of equalities (12.54) and (12.56), we will find that

$$A = \frac{(\alpha_1 + \alpha_2)^2}{2} - \left(\frac{\alpha_1^2}{3} + \alpha_2^2 + \alpha_1 + \alpha_2 \sqrt{1 + 2\alpha_1 + \frac{2}{3}\alpha_1^2 + \alpha_2^2}\right).$$

Substituting this expression into (12.56), we obtain the formula

$$\frac{Q}{N_0}=\frac{(\alpha_1+\alpha_2)^2}{2}-\alpha_1-\alpha_2\sqrt{1+2\alpha_1+\frac{2}{3}\alpha_1^2+\alpha_2^2+\sqrt{n}\left(\alpha_1+\alpha_2\sqrt{1+\frac{2\alpha_1}{\sqrt{n}}+\frac{\frac{2}{3}\alpha_1^2+\alpha_2^2}{n}}\right)}, \tag{12.57}$$

which we will called the adjusted Kaplan formula.

It is possible to use this formula for any given values of n [if condition /193 (12.55) is met].

Where n = 1 and where $n \gg 1$, formula (12.57) provides essentially precise results, while when the values of n are intermediate, the error a fortiori is less than 2 dB.

The family of curves in Figure 12.3 was plotted from formulas (12.57) and (12.45) for two error probability values. Depicted in Figure 12.4 for comparison

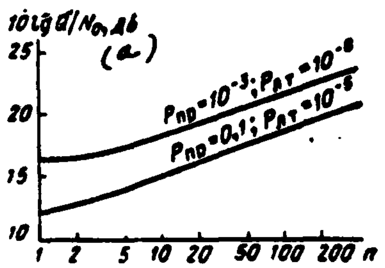


Figure 12.3. (a) — db.

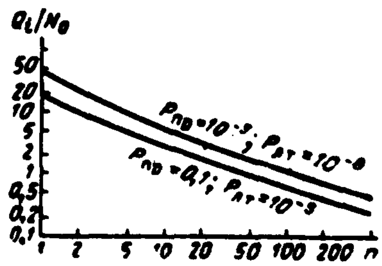


Figure 12.4

is dependence Q_i/N_0 on n, where

$$\frac{Q_i}{N_0}=\frac{1}{n}\cdot\frac{Q}{N_0}. \tag{12.58}$$

It is evident from Figures 12.3 and 12.4 that an increase in the number of pulses n in the packet means a decrease in requisite energy Q_i of each pulse, but packet energy Q as a whole increases. This denotes that division of the analog signal into a train of incoherent pulses is disadvantagous from the power standpoint.

The resultant loss in requisite energy is characterized by the ratio

$$\eta_1 = \frac{Q}{Q'}, \tag{12.59}$$

where Q' -- energy required for an analog (monopulse) signal, i. e., where n = 1.

It follows from (12.57) and (12.59) that

$$\eta_1 = 1 + \frac{2\sqrt{n}\left(\alpha_1 + \alpha_2\sqrt{1 + \frac{2\alpha_1}{\sqrt{n}} + \frac{\frac{2}{3}\alpha_1^2 + \alpha_2^2}{n}}\right) - 2\alpha_1 - 2\alpha_2\sqrt{1 + 2\alpha_1 + \frac{2}{3}\alpha_1^2 + \alpha_2^2}}{(\alpha_1 + \alpha_2)^2}. \tag{12.60}$$

Magnitude η_1 is called the energy loss due to pulse incoherence. If /194 pulses are coherent among themselves, but pulses are added without consideration of their rf occupation phases, i. e., with respect to law (12.17), then the same loss in energy occurs as is the case if the pulses themselves had, relative to each other, independent equiprobable random phases. Therefore, factor η_1 often is referred to also as the energy loss due to packet pulse adding incoherence.

The solid-line curves in Figure 12.5 were plotted from formula (12.60).

Since formula (12.60) was obtained from expression (12.57), then, just as is the case for (12.57), it only is precise when n = 1 and when $n \gg 1$. Where n values are intermediate, formula (12.60) provides somewhat-inflated magnitude η_1 values, however the error here a fortiori is less than 2 dB.

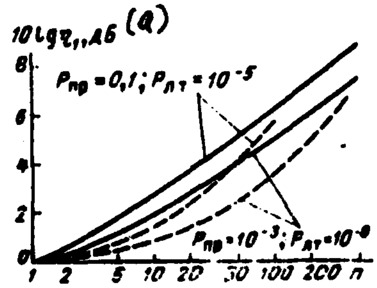


Figure 12.5. (a) -- dB.

The corresponding curves S. Ye. Fal'kovich [5] obtained is a somewhat more-precise way are plotted with dotted lines in Figure 12.5 for comparison. Evidently, the divergence among the corresponding curves here does not exceed 1.5 dB.

It follows from (12.60) that, when $n \gg 1$,

$$\eta_1 \approx \frac{2\sqrt{n}}{\alpha_1+\alpha_2} \approx \frac{2\sqrt{n}}{\sqrt{2\ln\frac{1}{P_{лт}}-2{,}8}+\sqrt{2\ln\frac{1}{P_{пр}}-2{,}8}} \qquad (12.61)$$

(here $P_{лт} \leqslant 0{,}1$ and $P_{пп} \leqslant 0{,}1$).

The following conclusion may be made based on formulas (12.60) and (12.61) and the Figure 12.5 curves:

1. The loss due to incoherence rises with a rise in n and, given large n values, this increase is proportional to $\sqrt{n}$.

2. The loss decreases with an increase in permissible error probabilities $P_{лт}$ and $P_{пр}$.

3. Where $n \leqslant 10$, the loss is relatively small and, given typical error probability values, does not exceed 1--3 dB.

On the other hand, given large n values, the loss may be quite considerable; therefore, if $n \gg 1$ and packet pulses are coherent among themselves, then it is very significant that their adding be coherent.

The aforementioned analysis applies to simple binary detection. The basic results and approach presented in the preceding chapter are fully applicable for analysis of simultaneous detection and recognition of m possible packets /195 $u_{k1}(t), \ldots, u_{km}(t)$. In particular, in the case of m orthogonal equally-probable packets with identical energies and given slight error probabilities ($P_{л\,т} \leqslant 0{,}1$ and $P_{пр} \leqslant 0{,}1$), requisite packet energy Q is determined by the simple binary detection formula if you replace $P_{л\,т}$ with $P_{л\,т}/m$ in it.

Therefore, for detection with recognition of m packets (or in the Figure 11.5 m-channel system), formulas (12.57), (12.60), and (12.61) in particular are valid if, in these formulas, α_1 is determined, not from formula (12.45), but the assumption is that:

$$\alpha_1^2 = 2\ln\frac{m}{P_{л\,т}} - 2{,}8 = 2\ln m + 2\ln\frac{1}{P_{л\,т}} - 2{,}8. \tag{12.62}$$

Here, for example, formula (12.61) takes the following form:

$$\eta_1 \approx \frac{2\sqrt{n}}{\sqrt{2\ln m + 2\ln\frac{1}{P_{л\,т}} - 2{,}8} + \sqrt{2\ln\frac{1}{P_{пр}} - 2{,}8}} \tag{12.63}$$

where $n \gg 1$, $P_{л\,т} < 0{,}1$, $P_{пр} < 0{,}1$.

It is evident from this formula that, when m increases, the loss due to incoherence decreases.

12.7 Detection and Recognition of Incoherent Packets With Amicably-Fluctuating Pulse Amplitudes

The following two extreme types of packet amplitude fluctuations are the most characteristic:

a) amicable fluctuations;

b) independent fluctuations.

For amicable fluctuations, the amplitude of pulses within a given packet is unchanged or is a known regular time function, while it changes from one packet to another (i. e., from one sequence to another) as an analog random magnitude with known a priori distribution.

For independent fluctuations, amplitudes of different packet pulses are independent random magnitudes.

We will examine amicable amplitude fluctuations in this section. Here, the second case examined in § 12.3 occurs here and the inverse probability of presence of a packet with amplitudes (a_1, . . ., a_n) is determined by expression (12.9):

$$P_y(a_1, \ldots, a_n) = P_y(a_1) = k_s P(a_1) e^{-\sum_{i=1}^{n} \frac{a_i^2 \tau_i}{2N_0}} \prod_{i=1}^{n} I_0\left(\frac{2a_i M_i}{N_0}\right), \tag{12.9}$$

where /196

$$a_i = a_1 + \Delta a_i;$$

Δa_i and τ_i -- precisely-known magnitudes:

$$P(a_1) = P(C)\, W(a_1); \tag{12.64}$$

here P(C) -- a priori probability of the presence of a packet with random amplitude a_1 value, while $W(a_1)$ -- law of distribution of packet pulse amplitude a_1, given that a packet is present.

It is convenient to represent expression (12.9) in the form

$$P_y(a_1) = k_s P(a_1) \exp\left\{-\sum_{i=1}^{n}\left[\frac{a_i^2 \tau_i}{2N_0} + \ln I_0\left(\frac{2a_i M_i}{N_0}\right)\right]\right\}. \tag{12.65}$$

In simple packet binary detection, a comparison is made of inverse probabilities

$P_y(C)$ and $P_y(0)$ of packet presence and absence, respectively. The decision is that a packet is present when this condition is met

$$P_y(C) > \eta P_y(0), \tag{12.66}$$

where η -- weight factor.

If condition (12.66) is not met, the decision is that no packet is present. Evidently,

$$P_y(C) = \int_0^\infty P_y(a_1)\, da_1; \tag{12.67}$$

therefore, considering (12.64) and (12.65), we have

$$P_y(C) = k_0 P(C) \int_0^\infty W(a_1) \exp\left\{-\sum_{i=1}^{n}\left[\frac{a_i^2\tau_i}{2N_0} + \ln I_0\left(\frac{2a_i M_i}{N_0}\right)\right]\right\} da_1. \tag{12.68}$$

If the a priori probability of packet absence equals P(0), then

$$P_y(0) = k_0 P(0). \tag{12.69}$$

Consequently, the decision must be made as to packet presence when this condition is met

$$y > U_0, \tag{12.70}$$

where

$$\left.\begin{aligned} y &= \int_0^\infty W(a_1) \exp\left\{-\sum_{i=1}^{n}\left[\frac{a_i^2\tau_i}{2N_0} + \ln I_0\left(\frac{2a_i M_i}{N_0}\right)\right]\right\} da_1; \\ a_i &= a_1 + \Delta a_i; \end{aligned}\right\} \tag{12.71}$$

$$U_0 = \eta \frac{P(0)}{P(C)}. \tag{12.72}$$

Further, for simplicity, we will assume that the amplitudes and durations /197 of all packet pulses are identical, i. e.,

$$a_1 = a_2 = \ldots = a; \quad \tau_1 = \tau_2 = \ldots = \tau;$$

then, expression (12.71) takes the form

$$y = \int_0^\infty W(a)\, e^{-\frac{na^2\tau}{2N_0}} e^{\sum_{l=1}^{n} \ln I_0\left(\frac{2a}{N_0} M_l\right)} da. \tag{12.73}$$

We will examine two extreme cases:

first case

$$\frac{Q_l}{N_0} \ll 1; \tag{12.74}$$

second case

$$\frac{Q_l}{N_0} \gg 1. \tag{12.75}$$

It was demonstrated in the preceding section that, in the first case, it is possible to assume:

$$\ln I_0\left(\frac{2aM_l}{N_0}\right) \approx \frac{1}{4}\left(\frac{2aM_l}{N_0}\right)^2,$$

while, in the second case

$$\ln I_0\left(\frac{2aM_l}{N_0}\right) \approx \frac{2aM_l}{N_0}.$$

Therefore, in the first case, from (12.73) we obtain

$$y=\int_0^\infty W(a)\,e^{-\frac{a^2\tau n}{2N_0}}\,e^{\frac{a^2}{2N_0^2}\sum_{i=1}^{n}M_i^2}\,da, \tag{12.76}$$

while, in the second case

$$y=\int_0^\infty W(a)\,e^{-\frac{a^2\tau n}{2N_0}}\,e^{\frac{2a}{N_0}\sum_{i=1}^{n}M_i}\,da. \tag{12.77}$$

It follows from (12.70) that, for packet detection, the receiver must compute magnitude y and compare it with threshold U_0. But, it follows from expressions (12.76) and (12.77) that, instead of comparision of magnitude y with threshold U_0, it is possible in the first case to compare with some other threshold U_0' magnitude

$$R_1=\sum_{i=1}^{n}M_i^2, \tag{12.78}$$

while, in the second case, magnitude /198

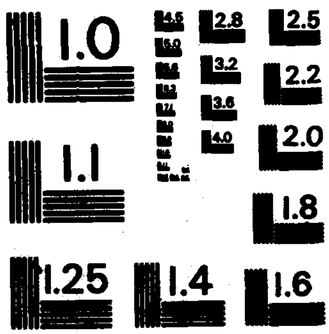
$$R_2=\sum_{i=1}^{n}M_i. \tag{12.79}$$

This possibility results from the fact that, as follows from (12.76) and (12.77), there is a regular monotonic dependence between y and R_1 (or y and R_2). Therefore, an overrun (or non-overrun) of some other threshold U_0' by magnitude R_1 (in the first case) or R_2 (in the second case) always will correspond to an overrun (or non-overrun) of threshold U_0 by magnitude y.

Expressions (12.78) and (12.79) coincide with the corresponding expressions obtained in the preceding section for packets witn known pulse amplitudes. Consequently, optimum receiver structure remains the same (Figure 12.2), with the only change being the magnitude of threshold bias U_0 at receiver output. Hence, it

AD-A120 899 THEORY OF OPTIMUM RADIO RECEPTION METHODS IN RANDOM NOISE(U) FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OH L S GUTKIN 24 SEP 82 FTD-ID(RS)T-0784-82 4/7

UNCLASSIFIED F/G 9/4 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

follows that a change in pulse amplitude law of distribution W(a) leaves the optimum receiver structure and mode unchanged, with the exception of the threshold bias U_0 magnitude.

A square-law detector [formula (12.78)] is optimum for every pulse, given a low power ratio Q_i/N_0, while a linear detector [formula (12.79)] is optimum, given a high ratio Q_i/N_0.

If the number n of packet pulses is small, then a high power ratio Q_i/N_0 is required for highly-reliable detection (i. e., slight error probabilities $P_{лт}$ and $P_{нр}$); if number n is very large, the reliable detection results already when $Q_i/N_0 \ll 1$ since power ratio nQ_i/N_0 turns out to be sufficiently-high for the packet as a whole. Therefore, in practice, the first case (12.74) occurs where $n \gg 1$, while the second case (12.75) occurs at slight n values.

The law of distribution of sums R_1 and R_2, determined from formulas (12.78) and (12.79), should be found in order to find error probabilities $P_{лт}$ and $P_{нр}$.

Since optimum receiver structure in this case is the same as for reception of packets with known amplitude, as examined in the preceding section, then, the law of distribution of magnitudes R_1 and R_2 turns out to be the same when there is no signal and, consequently, the false-alarm probability $P_{лт}$ expression remains unchanged. However, when there is a signal, the law of distribution of sums R_1 and R_2 is different than is the case for packets with known amplitude and will depend on the law of packet amplitude distribution W(a).

Kaplan analyzed the case of a Rayleigh law of distribution W(a) for the first time [16]. Kaplan only examined the first case, when

$$\frac{Q_i}{N_0} \ll 1 \quad \text{and} \quad n \gg 1.$$

His results using our designations may be reduced to the following approximate form: /199

$$\left.\begin{aligned}&\frac{Q_{cpi}}{N_0}\approx\frac{1}{P_{np}}\left(\frac{\alpha_1}{\sqrt{n}}+\frac{2+\alpha_1^2}{3n}\right);\\&\frac{Q_{cp}}{N_0}\approx\frac{1}{P_{np}}\left(\alpha_1\sqrt{n}+\frac{2+\alpha_1^2}{3}\right),\\&\alpha_1\approx\sqrt{2\ln\frac{1}{P_{\text{лт}}}-2,8.}\end{aligned}\right\}\qquad(12.80)$$

where

In these ratios, $Q_{cp}=nQ_{cpi}=n\frac{\overline{a^2}}{2}\tau$ is packet energy mean statistical value.

Formulas (12.80) are sufficiently precise for the following assumptions:

$$\frac{Q_i}{N_0}\ll 1;\ n\gg 1;\ P_{\text{лт}}\leqslant 0,1;\ P_{np}\leqslant 0,1. \qquad (12.81)$$

The error in formulas (12.80) asymptotically will strive towards zero due to strengthening of inequalities (12.81).

It follows from (12.80) that, given sufficiently-large values of n, it is possible to assume:

$$\frac{Q_{cp}}{N_0}=\frac{\sqrt{2\ln\frac{1}{P_{\text{лт}}}-2,8}}{P_{np}}\sqrt{n}. \qquad (12.82)$$

In light of a decrease in the number n of pulses, relative formula (12.80) error rises and reaches its greatest magnitude where n = 1. Here, from (12.80) we obtain:

$$\frac{Q_{cp}}{N_0}=\frac{1}{P_{np}}\left(\alpha_1+\frac{2+\alpha_1^2}{3}\right). \qquad (12.83)$$

where

$$\alpha_1 = \sqrt{2 \ln \frac{1}{P_{л\,т}} - 2{,}8}.$$

In actuality, where n = 1, relationship (9.84) occurs, i. e.,

$$\frac{Q_{ср}}{N_0} \approx \frac{1}{P_{пр}} \ln \frac{1}{P_{л\,т}}. \qquad (9.84)$$

It follows from comparison of relationships (12.83) and (9.84) that, when n = 1 (and $P_{л\,т} \leqslant 0{,}1$), the formula (12.80) error does not exceed 2 dB. Therefore, it is permissible to adjust formula (12.80) so that it will provide a precise result not only when n $\gg$ 1, but also when n = 1.

Using a procedure identical to that used in the preceding section, /200
i. e., instead of (12.80), assuming that

$$\frac{Q_{ср}}{N_0} = \frac{1}{P_{пр}} \left(\alpha_1 \sqrt{n} + \frac{2 + \alpha_1^2}{3} + A \right),$$

where A -- adjusting term not depending on n, and accepting

$$\alpha_1 \approx \sqrt{2 \ln \frac{1}{P_{л\,т}}},$$

we will obtain:

$$A = \frac{\alpha_1^2}{6} - \frac{2}{3} - \alpha_1;$$

$$\frac{Q_{ср}}{N_0} = \frac{1}{P_{пр}} \left[\ln \frac{1}{P_{л\,т}} + (\sqrt{n} - 1) \sqrt{2 \ln \frac{1}{P_{л\,т}}} \right]. \qquad (12.84)$$

We will call this formula an adjusted Kaplan formula for amicably-fluctuating packets. It provides a sufficiently-precise result if n = 1 and n $\gg$ 1 (where

$P_{\text{пр}} \leqslant 0{,}1$ and $P_{\text{л т}} \leqslant 0{,}1$). For intermediate values of n, formula (12.84) provides a somewhat-inflated ratio $Q_{\text{ср}}/N_0$ value. However, the resultant error here is less than 2 dB.

As was the case in the preceding section, we will introduce the concept of loss η_1 in requisite signal energy caused by packet pulse incoherence (or adding incoherence if the pulses are coherent):

$$\eta_1 = \frac{Q_{\text{ср}}}{Q'_{\text{ср}}}, \tag{12.85}$$

where $Q_{\text{ср}}'$ -- average energy required for an analog (monopulse) signal, i. e., where n = 1.

Then, from formula (12.84), we will obtain

$$\eta_1 = 1 + \frac{2(\sqrt{n} - 1)}{\sqrt{2 \ln \frac{1}{P_{\text{л т}}}}}. \tag{12.86}$$

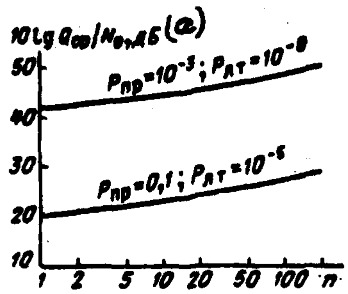


Figure 12.6. (a) -- dB.

The curves in Figures 12.6--12.8 are plotted from formulas (12.84) and (12.86).

It follows from comparison of the Figure 12.5 and 12.8 curves that the dependence of requisite signal energy on the number n of packet pulses in the case of amicably-fluctuating packets is approximately identical to that in the case of packets with a known amplitude. The loss due to packet pulse incoherence (or due /201

to pulse adding incoherence if the pulses are coherent) correspondingly are approximately identical as well.

The results presented above applied to a case of simple binary signal detection. The approach and basic results presented in the preceding chapter are applicable

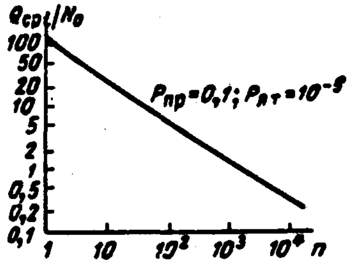


Figure 12.7

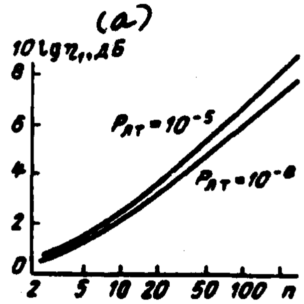


Figure 12.8. (a) — db

fully for simultaneous detection and recognition of m possible packets $u_{k1}(t)$, . . ., $u_{km}(t)$. In particular, in the case of m orthogonal equiprobable packets with identical average energies and slight error probabilities ($P_{лт} \leqslant 0.1$ and $P_{np} \leqslant 0.1$), requisite average packet energy Q_{cp} is determined from the simple binary detection formula (or curves) if $P_{лт}$ is replaced by $P_{лт}/m$ in it.

Here, formulas (12.84) and (12.86) take the following form (given the same assumptions):

$$\frac{Q_{cp}}{N_0} \approx \frac{1}{P_{np}}\left[\ln\frac{1}{P_{лт}}+\ln m+(\sqrt{n}-1)\sqrt{2\ln\frac{1}{P_{лт}}+2\ln m}\right]; \qquad (12.87)$$

$$\eta_1 = 1+\frac{2(\sqrt{n}-1)}{\sqrt{2\ln\frac{1}{P_{лт}}+2\ln m}}. \qquad (12.88)$$

12.8 Detection and Recognition of Incoherent Packets With Independently-Fluctuating Pulse Amplitudes

The third case examined in § 12.3 occurs for independent packet pulse amplitude

fluctuations and the inverse probability of the packet with amplitudes (a_1, . . ., a_n) is determined from formula (12.11), to wit

$$P_y(a_1, \ldots, a_n) = k\prod_{i=1}^{n} P(a_i)\, e^{-a_i^2\tau_i/2N_0} I_0\left(\frac{2a_i M_i}{N_0}\right).$$

Therefore, inverse probability $P_y(C)$ of packet presence (with any pulse /202 amplitudes) equals

$$P_y(C) = k\int_0^\infty \cdots \int_0^\infty P_y(a_1, \ldots, a_n)\, da_1 \ldots da_n =$$
$$= k\prod_{i=1}^{n}\int_0^\infty P(a_i)\, e^{-a_i^2\tau_i/2N_0} I_0\left(\frac{2a_i M_i}{N_0}\right) da_i. \quad (12.89)$$

But

$$P(a_i) = P(C)\, W(a_i), \quad (12.90)$$

where P(C) — a priori packet presence probability, and

$$P_y(C) = kP(C)\prod_{i=1}^{n}\int_0^\infty W(a_i)\, e^{-a_i^2\tau_i/2N_0} I_0\left(\frac{2a_i M_i}{N_0}\right) da_i. \quad (12.91)$$

Further, we will assume that amplitudes a_i fluctuate in accordance with the Rayleigh law. Then, comparing (12.91) with (9.56) and (9.65), we obtain:

$$P_y(C) = kP(C)\prod_{i=1}^{n}\frac{b_i}{\overline{u_{ci}^2}}\, e^{(2b_i/N_0)M_i^2}, \quad (12.92)$$

where, in accordance with (9.63) and (9.64)

$$b_i = \frac{\overline{u_{ci}^2}}{1+\frac{Q_{cp\,i}}{N_0}}; \quad Q_{cp\,i} = \frac{\overline{a_i^2}\tau_i}{2} = \overline{u_{ci}^2}\tau_i. \quad (12.93)$$

It follows from (12.92) that the inverse probability of packet absence (i. e., presence of a packet with zero pulse amplitudes) equals

$$P_y(0) = kP(0);$$

therefore, for simple binary detection, the decision as to signal (packet) presence must be made when this inequality is satisfied

$$P_y(C) > \eta P(0). \tag{12.94}$$

In future, we will accept for simplicity that all packet pulses have identical average energies and durations, i. e.,

$$Q_{cp\,1} = Q_{cp\,2} = \ldots = Q_{cp\,n};$$
$$\tau_1 = \tau_2 = \ldots = \tau;$$
$$Q_{cp} = nQ_{cp\,1};$$

then, considering relationships (12.92) and (12.93), it is possible to write signal presence condition (12.94) in the following form:

$$y > U_0, \tag{12.95}$$

where /203

$$y = \sum_{i=1}^{n} M_i^2; \tag{12.96}$$

$$U_0 = \frac{N_0 \tau}{2} \left(1 + \frac{N_0 n}{Q_{cp}}\right) \left[\ln\left(\eta \frac{P(0)}{P(C)}\right) + \ln\left(1 + \frac{Q_{cp}}{nN_0}\right)\right]. \tag{12.97}$$

Here, as was the case earlier [see (12.9a)],

$$M_i^2 = X_i^2 + Y_i^2;$$

therefore

$$y=\sum_{i=1}^{n}(X_i^2+Y_i^2). \tag{12.98}$$

It follows from (12.96) that, in the case being examined, i. e., for independent packet pulse fluctuations, a square-law characteristic is the optimum response curve.

Initially, we will examine a case where there is no signal in order to find detection error probability.

It follows from (12.9a) that, when there is no signal

$$\begin{aligned} X_i &= \int_{t_i}^{t_i+\tau} u_{\text{m}}(t)\cos\omega t\,dt; \\ Y_i &= \int_{t_i}^{t_i+\tau} u_{\text{m}}(t)\sin\omega t\,dt. \end{aligned} \tag{12.99}$$

It was demonstrated in § 5.2 that random magnitude $\frac{2}{N_0}\int_0^T u_{\text{m}}(t)u_{\text{c}}(t)dt$ has a normal law of distribution with zero mean value and dispersion $2Q/N_0$, where $Q=\int_0^T u_{\text{c}}^2(t)\,dt$; hence, it follows that magnitudes X_i and Y_i, determined by expressions (12.99), have a normal distribution with zero mean values and dispersions σ_i^2,

$$\sigma_i^2=\frac{\left(\frac{2Q}{N_0}\right)}{\left(\frac{2}{N_0}\right)^2}=\frac{\left(\frac{2\cdot 1^2\cdot\tau}{2N_0}\right)}{\left(\frac{2}{N_0}\right)^2}=\frac{N_0\tau}{4}. \tag{12.100}$$

Magnitudes X_i and Y_i mutually are statistically independent due to the

orthogonality of functions cos ω t and sin ω t, included in expression (12.99), and a normal law of noise distribution.

Magnitudes X_i and Y_i with varied numbers i also are statistically independent since the integration intervals corresponding to them in expressions (12.99) do not overlap. Thus, magnitude y, determined from expression (12.98), is the /204 sum of the squares 2n of independent normally-distributed random magnitudes with zero mean values and identical dispersions σ_l^2 .

Therefore, when σ_l^2 = 1, magnitude y has law of distribution χ^2_{2n} (chi-square, with 2n degrees of freedom). This law of distribution has been tabularized (see the attachment, for example); meanwhile, magnitude $\chi^2_{2n}(\alpha)$, where $\chi^2_{2n}(\alpha)$ is that magnitude, the probability of overrun of which is the total magnitude y (where σ_l^2 = 1) equals α , i. e., where σ_l^2 = 1, usually is presented in tables

$$P\left[y>\chi^2_{2n}(\alpha)\right]=\alpha. \qquad (12.101)$$

In our case, dispersion σ_l^2 does not equal unity and is determined from expression (12.100). Therefore, instead of (12.101), we should assume:

$$P\left[y>\sigma_l^2\chi^2_{2n}(\alpha)\right]=\alpha. \qquad (12.102)$$

It follows from (12.95) that the probability of satisfying inequality (12.95) when there is no signal is the false-alarm probability, i. e., when there is no signal

$$P(y>U_0)=P_{\pi\tau}. \qquad (12.103)$$

It follows from comparison of (12.102) and (12.103) that

$$U_0=\sigma_l^2\,\chi^2_{2n}(P_{\pi\tau}), \qquad (12.104)$$

or, considering (12.100)

$$\chi^2_{2n}(P_{\pi\tau})=2\left(1+\frac{N_n n}{Q_{cp}}\right)\left[\ln\left(\eta\frac{P(0)}{P(C)}\right)+\ln\left(1+\frac{Q_{cp}}{nN_0}\right)\right]. \qquad (12.105)$$

It is possible to compute $\chi^2_{2n}(P_{\pi\tau})$ from this formula for given magnitudes $\eta P(0)/P(C)$, Q_{op}/N_0 and n, and then, using the $\chi^2_{2n}(\alpha)$ tables of distribution, to find the corresponding value of the argument, i. e., false-alarm probability $P_{\pi\tau}$.

We now will examine the case where there is a signal.

When there is a signal, the probability that inequality (12.95) is satisfied is correct signal detection probability $P_{\pi o}$, i. e.,

$$P(y > U_n) = P_{\pi o} = 1 - P_{np}, \tag{12.106}$$

where magnitude y, determined from relationship (12.96), is now signal-plus-noise (pulse packets).

We will assume for specificity that signal pulse amplitude is distributed in accordance with the Rayleigh law. The phase of these pulses was accepted above as random and equally probable. Thus, this signal has, for each pulse, a random equiprobable phase and random amplitude distributed in accordance with the Rayleigh law. Here, amplitude and phase values are statistically independent for various packet pulses. This signifies that instantaneous signal voltage values at moments $t_i + \tau$ have a normal law of distribution, i. e., the same law of distribution that the noise has. Here, signal-plus-noise also has a normal law of distribution.

Therefore, when there is a signal, the law of distribution of terms /205 X_i and Y_i in expression (12.98) and, consequently, the magnitude y law of distribution as well, remain identical to that when only noise is present. However, dispersion σ_i^2 is determined now, not from expression (12.100), but increases by a factor of μ^2 due to signal action where, in accordance with (9.78)

$$\mu^2 = 1 + \frac{Q_{op\,i}}{N_0}.$$

Therefore, when there is a signal, instead of (12.100), this expression occurs

$$\sigma_i^2=\frac{N_0\tau}{4}\mu^2=\frac{N_0\tau}{4}\left(1+\frac{Q_{cp}\,t}{N_0}\right). \tag{12.107}$$

Since, when there is a signal, magnitude y has the identical distribution as is the case when there is no signal, then formula (12.102) remains valid.

On the other hand, as noted above, when there is a signal

$$P(y>U_0)=P_{\text{п о}}.$$

When this expression is compared with (12.102), we have

$$\sigma_i^2\chi_{2n}^2(P_{\text{п о}})=U_0.$$

Considering (12.97) and (12.107), finally we obtain:

$$\chi_{2n}^2(P_{\text{п о}})=\frac{2\left(1+\frac{N_0 n}{Q_{cp}}\right)}{\left(1+\frac{Q_{cp}}{N_0 n}\right)}\left[\ln\left(\eta\frac{P(0)}{P(C)}\right)+\ln\left(1+\frac{Q_{cp}}{N_0 n}\right)\right]. \tag{12.108}$$

It is possible to compute $\chi_{2n}^2(P_{\text{п о}})$ from this formula for given $\eta P(0)/P(C)$, Q_{cp}/N_0 and n and then use the $\chi_{2n}^2(\alpha)$ table of distribution to find the corresponding values of the argument, i. e., magnitude $P_{\text{п о}}$.

If a priori probabilities P(0) and P(C) are unknown, then magnitudes $P_{\text{л т}}$ and $P_{\text{п о}}$ must be given. Here, as follows from (12.107) and (12.108), the following relationship determines requisite signal energy

$$\frac{Q_{cp}}{N_0}=n\left[\frac{\chi_{2n}^2(P_{\text{л т}})}{\chi_{2n}^2(P_{\text{п о}})}-1\right]. \tag{12.109}$$

Using the $\chi^2_{2n}(\alpha)$ tables of distribution, it is not difficult using formula (12.109) to compute signal-to-noise Q_{cp}/N_0 required to provide given probabilities $P_{\text{лт}}$ and P_{np} for given number of pulses n.

Distribution $\chi^2_{2n}(\alpha)$, even in the most complete tables, is given only for $n \leqslant 50$. However, for greater values of n, it is possible to approximate the magnitude y distribution [formula (12.96)] using a normal distribution that /206 is similar to what was done in § 12.6. Then, instead of (12.109), we obtain relationship

$$\frac{Q_{cp}}{N_0} \approx (\alpha_1 + \alpha_2)\sqrt{n}, \quad \text{where} \quad n \gg 1. \tag{12.110}$$

This expression coincides with the corresponding expression (12.48) obtained in § 12.6 for a signal with a known amplitude, with the exception that, instead of signal energy Q, it contains the mean statistical value of this energy Q_{cp}. As usual, magnitudes α_1 and α_2 are determined from formulas (12.45).

For large n values, Kaplan obtained a slightly-more precise expression [16], approximating distribution χ^2_{2n}, not by a normal distribution, but with an expansion into a power series with respect to a normally-distributed magnitude. Here, instead of (12.110), the following more-precise expression is obtained:

$$\frac{Q_{cp}}{N_0} \approx (\alpha_1 + \alpha_2)\left[\sqrt{n} + \frac{\alpha_1 + 2\alpha_2}{3}\right]. \tag{12.111}$$

where α_1 and α_2 are the usual values.

Given sufficiently-large values of n, expression (12.111) coincides with (12.110), as consequently would be expected.

Formula (12.111) is sufficiently precise already when

$$n > 15 - 20.$$

The curves depicted in Figure 12.9 are plotted from formulas (12.109) and (12.111).

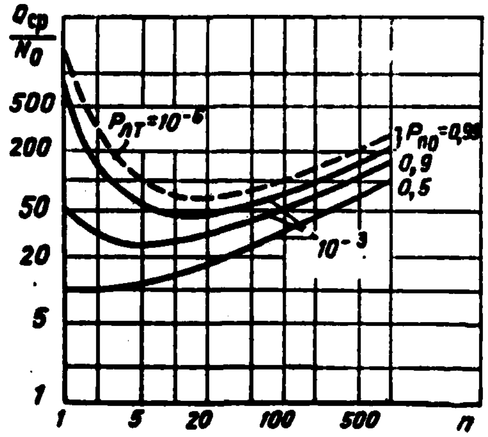


Figure 12.9

We now will find loss η_1, determined from relationship (12.85), in which Q_{cp}' is value Q_{cp} occurring when n = 1 and determined from formula (9.83).

It follows from formulas (12.109) and (9.83) that

$$\eta_1 = \frac{n\left[\frac{\chi^2_{2n}(P_{лт})}{\chi^2_{2n}(P_{по})} - 1\right]}{\left(\frac{\ln\frac{1}{P_{лт}}}{\ln\frac{1}{P_{по}}} - 1\right)}. \tag{12.112}$$

Where $n \geqslant 15 - 20$, in accordance with formulas (9.83) and (12.111), we obtain

$$\eta_1 \approx \frac{(\alpha_1 + \alpha_2)\sqrt{n} + \frac{\alpha_1 + 2\alpha_2}{3}}{\left(\frac{\ln\frac{1}{P_{лт}}}{\ln\frac{1}{P_{по}}} - 1\right)}; \tag{12.113}$$

where $P_{лт} \leqslant 0{,}1$ and $P_{пр} \leqslant 0{,}1$, it is possible, in accordance with (12.45), /207
to assume

$$\begin{aligned} \alpha_1 &\approx \sqrt{2\ln\frac{1}{P_{лт}} - 2{,}8}; \\ \alpha_2 &\approx \sqrt{2\ln\frac{1}{P_{пр}} - 2{,}8}. \end{aligned} \tag{12.45}$$

Finally, if $n \gg 1$, $P_{лт} \ll 1$ and $P_{пр} \ll 1$, from (12.113) and (12.45), we obtain

$$\eta_1 \approx \frac{P_{пр}\left(\sqrt{2\ln\frac{1}{P_{лт}}} + \sqrt{2\ln\frac{1}{P_{пр}}}\right)\sqrt{n}}{\ln\frac{1}{P_{лт}}}. \tag{12.114}$$

It is possible to plot the curves depicted in Figure 12.10 from formulas (12.109)--(12.114) or from the Figure 12.9 curves.

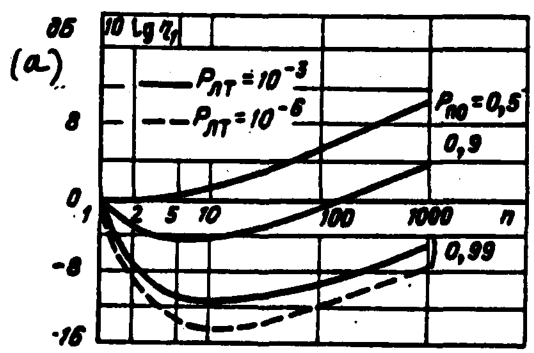


Figure 12.10. (a) -- db.

It is possible to draw the following basic conclusions from analysis of the formulas presented above and the Figure 12.9 and 12.10 curves:

1. In the case of independently-fluctuating pulses, dependence of requisite signal energy on the number of pulses n has (where $P_{no} \geqslant 0.5$) a minimum, given some optimum number of pulses n_{onT} .

2. Magnitude n_{onT} will depend little on permissible false-alarm probability $P_{\pi T}$, but will depend greatly on permissible correct detection probability P_{no} , increasing as probability P_{no} rises. Thus, for example, where P_{no} = 0.9, the result is $n_{onT} \approx 5$; if P_{no} = 0.99, then $n_{onT} \approx 10$.

3. The gain occurring when n = n_{onT} , in comparison with n = 1, rises with a decrease in error probabilities P_{np} and $P_{\pi T}$ and will depend especially radically on probability P_{np}. Thus, for example, if where P_{np} = 0.1 and $P_{\pi T} = 10^{-3}$, /208 maximum gain equals 4.2 dB, with this increasing already to 15 dB where P_{np} = 0.01 and $P_{\pi T} = 10^{-6}$.

4. It follows from comparison of relationships (12.48) and (12.110) that, when $n \gg 1$, the average energy required for detection of a packet with independently-fluctuating amplitudes equals the energy required for identically-reliable detection of a packet with known amplitudes.

If n = 1, then, in the case of a fluctuating amplitude, considerably-greater signal energy is required for reliable detection ($P_{\pi T} \leqslant 0.1$ and $P_{np} \leqslant 0.1$) than is the case for known amplitude.

It follows from the aforementioned points that, given high requirements levied for fluctuating signal detection reliablity (given slight P_{np} and $P_{\pi T}$), it is possible to obtain a significant gain in requisite signal energy if an analog (monopulse) signal is divided into n sufficiently-separated over time (or frequency occupation) pulses so that the fluctuations of adjacent pulses may be considered statistically independent. Here, it is desirable to select $n \approx n_{onT}$.

The possibility of obtaining a gain in requisite energy occurs here because the probability that all packet pulses n will "fade" (decrease in amplitude) at

the same time is significantly lower than the probability of such complete fading of a monopulse signal.

The above analysis applied to a case of simple binary detection. The approach and basic results presented in the preceding chapter are applicable for simultaneous detection and recognition of m possible packets (or in the case of an m-channel receiving system).

In particular, in the case of m orthogonal equiprobable packets with identical mean values and given slight error probabilities ($P_{лт} \leqslant 0{,}1$ and $P_{пр} \leqslant 0{,}1$), requisite packet average energy $Q_{ср}$ is determined from simple binary detection formulas (or curves) if $P_{лт}$ is replaced by $P_{лт}/m$.

Thus, for example, formulas (12.109) and (12.111) take the form /209

$$\frac{Q_{ср}}{N_0} = n\left[\frac{\chi^2_{2n}\left(\frac{P_{лт}}{m}\right)}{\chi^2_{2n}(P_{по})} - 1\right]. \qquad (12.115)$$

Where n $\geqslant$ 15 – 20

$$\frac{Q_{ср}}{N_0} \approx (\alpha_1 + \alpha_2)\left(\sqrt{n} + \frac{\alpha_1 + 2\alpha_2}{3}\right). \qquad (12.116)$$

where (for $P_{лт} \leqslant 0{,}1$ and $P_{пр} \leqslant 0{,}1$))

$$\alpha_1 \approx \sqrt{2\ln\frac{1}{P_{лт}} + 2\ln m - 2{,}8}; \quad \alpha_2 \approx \sqrt{2\ln\frac{1}{P_{пр}} - 2{,}8}.$$

12.9 Special Features of Signal Detection and Recognition in Presence of Nonwhite Noise

The entire analysis presented above for signal detection and recognition involved noise in the form of normal white noise. S. Ye. Fal'kovich [5] for the

first time obtained corresponding results for normal nonwhite noise using the approach presented in Chapter 3 for reducing nonwhite noise to white noise.

Let $S(j\omega)$ and $S_{\text{ш}}(\omega)$ be signal $u_c(t)$ complex spectrum and noise $u_{\text{ш}}(t)$ power spectrum, respectively. Then, the optimum receiver will comprise (Figure 3.1) correcting linear two-port ("whitening" filter) with transfer constant $K_1(j\omega)$ and receiver Π' optimum for white noise $u_{\text{ш}}'(t)$ with spectral energy distribution (unilateral) N_0 and transformed signal $u_c'(t)$ with spectrum $S'(j\omega)$.* Here

$$|K_1(j\omega)|^2=\frac{b}{S_{\text{ш}}(\omega)}, \tag{12.117}$$

where b — some constant:

$$S'(j\omega)=K_1(j\omega)S(j\omega). \tag{12.118}$$

Here

$$N_0=2|K_1(j\omega)|^2S_{\text{ш}}(\omega)=2b. \tag{12.119}$$

We will examine the problem of the detection with recognition of m signals $u_1(t), \ldots, u_k(t), \ldots, u_m(t)$ having spectrum $S_1(j\omega), \ldots, S_k(j\omega), \ldots, S_m(j\omega)$, on a background of nonwhite noise $u_{\text{ш}}(t)$ with power spectrum $S_{\text{ш}}(\omega)$. Then, due to what has been presented above, the frequency characteristic of the "whitening" filter is determined from relationship (12.117) and the task will boil down to finding receiver Π', which optimally will solve the problem of detection with recognition of m transformed signals $u_1'(t), \ldots, u_k'(t), \ldots, u_m'(t)$ on the background of white noise $u_{\text{ш}}'(t)$ with spectral energy distribution N_0.

Here, in accordance with (12.117) and (12.118), the spectrum of the k-th /210 transformed signal (k = 1, . . ., m) equals

$$S_k'(j\omega)=\frac{b}{S_{\text{ш}}(\omega)}S_k(j\omega). \tag{12.120}$$

*See note on page 54.

A priori probabilities of the appearance of transformed signals $u_k'(t)$ $(k = 1, \ldots, m)$ coincide with the a priori probabilities of corresponding initial signals $u_k(t)$ $(k = 1, \ldots, m)$. Therefore, a priori (inverse) probabilities $P_y(u_o)$, $P_y(u_1)$, . . ., $P_y(u_m)$ may be found and compared by using the optimum system, the circuitry of which is depicted in Figure 11.1, described in the preceding section. Here, y(t) should be understood to mean signal-plus-noise at <u>output</u> of the "whitening" filter with transfer function (12.117). Since noise $u_{ш}'(t)$ included in signal-plus-noise y(t) turns out here to be white, the approaches described in Chapter 11 relative to white noise may be used to find the structure of receivers $П_k$ $(k = 1, \ldots, m)$.

In particular, receiver $П_k$ input stages may be realized in the form of linear filters matched with transformed signals $u_k'(t)$. Here, the transfer constant of the k-th matched filter equals

$$K_k(j\omega) = aS_k'^*(j\omega)\,e^{-j\omega t_0},$$

where a -- some constant;

$$S_k'^*(j\omega) = S_k'(-j\omega).$$

For simplicity, we will limit ourselves to examination of the following case for computation of error probabilities and determination of the requisite signal-to-noise ratio.

We will assume that input signals $u_1(t)$, . . ., $u_m(t)$ are equally probable, orthogonal, and have identical energy Q. In addition, we will assume spectra $S_k(j\omega)$ $(k = 1, \ldots, m)$ and $S_ш(\omega)$ are such that transformed signals $u_1'(t)$, . . ., $u_m'(t)$ retain the property of orthogonality and equality of energies (this occurs, for instance, if amplitude spectra $|S_k(j\omega)|$ $(k = 1, \ldots, m)$ of all signals are identical, while signal orthogonality insures a sufficiently-great separation of their activities over time).

When these conditions are met, the task will boil down to detection with recognition of m equiprobable orthogonal signals $u_1'(t)$, . . ., $u_m'(t)$ with identical

where

$$\mu = \frac{\int\limits_0^\infty |S(j2\pi f)|^2 df}{\int\limits_0^\infty \frac{|S(j2\pi f)|^2 df}{\left[\frac{G_m(f)}{G_m(f_0)}\right]}}. \tag{12.127}$$

Formula (12.126) provides the link between power ratio $Q/G_m(f_0)$, required in the case of nonwhite noise with spectrum $G_m(f)$, and corresponding ratio Q'/N_0 for white noise, which was determined in preceding chapters (in the case of fluctuating signals, Q and Q' are replaced by Q_{cp} and Q_{cp}').

Consequently factor μ demonstrates by what factor requisite signal-to-noise power ratio at system input changes when nonwhite noise replaces white noise.

Thus, for example, in a case of nonwhite noise, formula (12.87) takes the following form:

$$\frac{Q_{cp}}{G_m(f_0)} \approx \mu \frac{1}{P_{np}}\left[\ln\frac{1}{P_{\pi\tau}} + \ln m + (\sqrt{n}-1)\sqrt{2\ln\frac{1}{P_{\pi\tau}} + 2\ln m}\right], \tag{12.128}$$

where μ is determined from formula (12.127).

As our illustration, we will examine the same example as was used in Chapter 3, thereby assuming that signal and noise spectra at system input have the form

$$\left.\begin{aligned} |S(j2\pi f)|^2 &= S_0^2 e^{-(f-f_0)^2/2\Pi_c^2}; \\ G_m(f) &= G_m(f_0) e^{-(f-f_0)^2/2\Pi_m^2}. \end{aligned}\right\} \tag{12.129}$$

Substitution of these relationships into formulas (12.126) and (12.127) provides

$$\frac{Q}{G_{\text{ш}}(f_0)}=\frac{Q'}{N_0}\cdot\frac{1}{\sqrt{1-\left(\frac{\Pi_{\text{c}}}{\Pi_{\text{ш}}}\right)^2}}. \tag{12.130}$$

It was explained in Chapter 3 that the optimum system for type (12.129) signal and noise spectra turns out to approximate the physically-realizable system only when condition (3.12) is met.

Therefore, formula (12.130) makes sense also only when this condition /213 is met

$$\frac{\Pi_{\text{c}}}{\Pi_{\text{ш}}}\leqslant\frac{1}{\sqrt{2}}. \tag{12.131}$$

It follows from formula (12.130) that $\Pi_{\text{ш}}/\Pi_{\text{c}}\geqslant 2-3$ results when $Q/G_{\text{ш}}(f_0) \approx Q'/N_0$.

This denotes that, if the width of the noise power spectrum is even greater by a factor of 2—3 than that of the signal spectrum, then such noise essentially acts just like white noise during detection and recognition of orthogonal signals.

CHAPTER THIRTEEN /213

MEASUREMENT OF SIGNAL ANALOG PARAMETERS

13.1 Measurement of Random Initial Phase Signal Amplitude

Let

$$y(t)=u_{a,\varphi}(t)+u_{\text{m}}(t), \tag{13.1}$$

where

$$u_{a,\varphi}(t)=a\cos(\omega t+\varphi) \tag{13.2}$$

is a signal, the amplitude a of which requires measurement. Here, frequency ω is assumed known, while initial phase φ is random and equally probable, i. e., has distribution (9.3).

Amplitude a has known a priori distribution P(a). Amplitude a and phase φ are assumed to be analog random magnitudes, i. e., they are constant during observation cycle (0, T), while they change from one cycle to another with indicated probability densities P(a) and P(φ), respectively.

The amplitude measurement problem will comprise, evidently, solution, based

on analysis of realization y(t), of which value amplitude a will have during a given observation cycle (0, T). In accordance with the inverse probability approach presented in Chapter 8, this requires computation of distribution $P_y(a)$.

For the case being examined, distribution $P_y(a)$ was computed in § 9.1 and is determined from formula (9.15). The problem of the structure of the optimum receiver computing distribution (9.15) or (9.15a) was examined in the same section.

Especially-simple and clear results are obtained if one assumes that a priori distribution P(a) is uniform, while the signal-to-noise ratio is sufficiently high. Therefore, we will examine this case in more detail.

For an equidimensional a priori amplitude distribution, formula (9.15) /214 takes the form

$$P_y(a) = k_3 e^{-a^2T/2N_0} I_0\left(\frac{2aM}{N_0}\right), \tag{13.3}$$

where k_3 -- constant normalizing factor.

Assuming

$$\frac{dP_y(a)}{da} = 0,$$

we will find the most-probable value $a_{y\,n}$ of amplitude a corresponding to the curve $P_y(a)$ curve:

$$a_{y\,n} = \frac{2M}{T} \cdot \frac{I_1\left(\frac{2a_{y\,n}M}{N_0}\right)}{I_0\left(\frac{2a_{y\,n}M}{N_0}\right)}, \tag{13.4}$$

where $I_0(z)$ and $I_1(z)$ -- modified Bessel functions (imaginary argument jz).

Equation (13.4) is transcendental; however, given a high signal-to-noise ratio, it is possible to assume, considering the properties of Bessel functions $I_0(z)$ and $I_1(z)$, that

$$\frac{I_1\left(\frac{2a_{yn}M}{N_0}\right)}{I_0\left(\frac{2a_{yn}M}{N_0}\right)} \approx 1$$

and

$$a_{yn} = \frac{2}{T} M. \quad (13.5)$$

In addition, given a high signal-to-noise ratio, it is possible to assume

$$I_0\left(\frac{2aM}{N_0}\right) \approx \frac{e^{2aM/N_0}}{\sqrt{2\pi \frac{2aM}{N_0}}}, \quad (13.6)$$

since, where $z \geqslant 3$

$$I_0(z) = \frac{e^z}{\sqrt{2\pi z}}. \quad (13.7)$$

Considering relationships (13.5) and (13.6) and normality condition

$$\int_0^\infty P_y(a)\, da = 1,$$

it is possible to reduce expression (13.3) to the following form, valid for a high signal-to-noise ratio:

$$P_y(a) = \sqrt{\frac{T}{2\pi N_0}}\, e^{-\frac{T}{2N_0}(a-a_{yn})^2}. \quad (13.8)$$

If amplitude a is modulated by message x in accordance with the law /215

$$a = u_0(1 + mx), \tag{13.9}$$

where magnitude u_0 is precisely known, then, replacing variable a with x in (13.8), we obtain

$$P_y(x) = \sqrt{\frac{b}{\pi}} e^{-b(x - x_{y\,n})^2}, \tag{13.10}$$

where $x_{y\,n}$ — most-probable message x value and

$$b = \frac{m^2 Q_0}{N_0}; \quad Q_0 = \frac{u_0^2 T}{2}. \tag{13.11}$$

Comparing formulas (13.10) and (13.11) with corresponding formulas (6.12) and (6.22) obtained in Part II of this book for a precisely-known signal, it is not difficult to become convinced that they are identical. Consequently, given a high signal-to-noise ratio, resultant distribution $P_y(x)$ for a random-phase AM signal is identical to that for a precisely-known AM signal.

The optimum receiver will reproduce message x exclusively on the basis of analysis of distribution $P_y(x)$. Therefore, noise immunity during reception of a random-phase AM signal is identical to that during reception of a precisely-known signal if the signal-to-noise ratio is high. Here, the error in message x reproduction is described by the same formulas (6.17) and (6.22).

Consequently, given a high signal-to-noise ratio, the random nature of signal rf occupation phase φ will not lead to message reproduction deterioration.

If the signal-to-noise ratio is not high, then the resultant analysis is considerably more unwieldy. It demonstates that the lower this ratio, the greater the deterioration in noise immunity phase irregularity will lead to.

It follows from formula (13.5) that, given a high signal-to-noise ratio,

output voltage γ of an optimum receiver operating on the maximum inverse probability principle, i. e., on the principle

$$\gamma = a_{y\,n},$$

is determined by very-simple relationship

$$\gamma = \text{const}\ M, \tag{13.12}$$

where M — envelope of the optimum linear filter output voltage (at moment T). Consequently, in this case, the optimum receiver will comprise an optimum linear filter and a linear amplitude detector (Figure 9.2).

It was demonstrated in § 2.4 that, if only one signal pulse arrives during each observation cycle (0, T), then the optimum linear filter may be replaced almost without loss by a conventional band-pass amplifier (for instance, by a conventional if amplifier) with optimum bandwidth. In this case, the optimum receiver boils down to a standard AM signal receiver comprising a band-pass /216 amplifier and linear amplitude detector.

As already noted above, the mean square of the error in normalized message x reproduction, given a high signal-to-noise ratio, as usual is determined from formulas (6.17) and (6.22), from which it folllows that

$$\overline{\delta^2} = \frac{N_0}{2Q_0 m^2}, \tag{13.13}$$

where

$$Q_0 = \frac{u_0^2 T}{2}$$

is relative signal energy where x = 0, while

$$\delta = x_{y\,n} - x. \tag{13.14}$$

We now will find the mean square of the amplitude measurement error, i. e., magnitude $(\overline{\Delta a^2})$, where

$$\Delta a = a_{y\,u} - a \tag{13.15}$$

(here, just as before, it is assumed that the optimum receiver operates on the maximum inverse probability density principle).

It follows from relationships (13.9), (13.14), and (13.15) that

$$\overline{(\Delta a)^2} = u_0^2 m^2 \overline{\delta^2} \tag{13.16}$$

or, considering formula (13.13)

$$\overline{(\Delta a)^2} = \frac{N_0}{T}, \tag{13.16a}$$

i. e., the mean square of the amplitude measurement error will depend only on relative noise energy per unity of band, N_0, and observation time T.

13.2 Measurement of Random-Phase Pulse Signal Moment τ of Arrival

Let

$$y(t) = a(t-\tau)\cos(\omega t + \varphi) + u_{ш}(t), \tag{13.17}$$

where $a(t - \tau)$ -- envelope of a precisely-known signal, with the exception of time shift τ, subject to measurement.

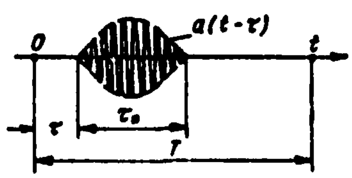


Figure 13.1

Since this is assumed to be a pulse signal (Figure 13.1), then

$$\left.\begin{array}{l} a(t-\tau)=0, \\ \text{if } t<\tau \text{ then } t>\tau+\tau_{u}, \end{array}\right\} \tag{13.18}$$

where τ_u -- signal pulse known duration.

Time of observation T encompasses the interval of all possible signal pulse positions corresponding to all possible measured shift τ values. /217

Frequency ω is assumed to be known, while initial phase -- is a parasitic random parameter distributed in accordance with law (9.3). Parameters τ and φ are assumed to be analog independent random magnitudes, i. e., they are constant during observation cycle (0, T).

Distribution $P_y(\tau)$ analysis must form the basis for measurement of moment τ. It follows from (8.10) and (13.17) that

$$P_{\tau,\varphi}(y)=\frac{1}{(\sqrt{2\pi N})^n}\exp\left\{-\frac{1}{N_0}\int_0^T[y(t)-a(t-\tau)\cos(\omega t+\varphi)]^2\,dt\right\}.$$

Considering (8.4), (8.6), and (8.8), we have

$$\begin{aligned} P_y(\tau)&=\int_0^{2\pi}P_y(\tau,\varphi)\,d\varphi=\int_0^{2\pi}kP(\tau)P(\varphi)P_{\tau,\varphi}(y)\,d\varphi=\\ &=k_1P(\tau)\int_0^{2\pi}e^{-\frac{1}{N_0}\int_0^T a^2(t-\tau)\cos^2(\omega t+\varphi)\,dt}\cdot e^{\eta(\tau,\varphi)}\,d\varphi, \end{aligned} \tag{13.19}$$

where $P(\tau)$ -- a priori parameter τ distribution,

$$\eta(\tau,\varphi)=\frac{2}{N_0}\int_0^T y(t)\,a(t-\tau)\cos(\omega t+\varphi)\,dt, \tag{13.20}$$

while k_1 -- normalizing constant.

Since interval (0, T) includes all possible signal pulse positions, then

$$\int_0^T a^2(t-\tau)\cos^2(\omega t+\varphi)\,dt = Q, \tag{13.21}$$

where Q will not depend on τ. Therefore, from (13.19) we have

$$P_y(\tau) = k_2 P(\tau)\int_0^{2\pi} e^{\eta(\tau,\varphi)}\,d\varphi, \tag{13.22}$$

where k_2 -- constant determined from a normality condition.

It follows from (13.18) and (13.20) that

$$\eta(\tau,\varphi) = \frac{2}{N_0}\int_{\tau}^{\tau+\tau_n} y(t)\,a(t-\tau)\cos(\omega t+\varphi)\,dt. \tag{13.23}$$

It is not difficult to reduce this expression to the following form: /218

$$\eta(\tau,\varphi) = \frac{2}{N_0} M_1(\tau)\cos(\theta+\varphi), \tag{13.24}$$

where

$$\left.\begin{aligned} M_1(\tau) &= \sqrt{X_1^2+Y_1^2};\\ X_1 &= \int_{\tau}^{\tau+\tau_n} y(t)\,a(t-\tau)\cos\omega t\,dt;\\ Y_1 &= \int_{\tau}^{\tau+\tau_n} y(t)\,a(t-\tau)\sin\omega t\,dt. \end{aligned}\right\} \tag{13.25}$$

Substituting (13.24) into (13.22), we obtain

$$P_y(\tau) = k_2 P(\tau)\, I_0\left[\frac{2M_1(\tau)}{N_0}\right]. \tag{13.26}$$

where k_3 -- constant determined from a normality condition.

It follows from (13.23) and (13.24) that $M_1(\tau)$ is the envelope with respect to φ of oscillation $\frac{N_0}{2}\eta(\tau, \varphi)$, i. e., of oscillation

$$\xi(\tau, \varphi) = \int_{\tau}^{\tau+\tau_{u}} y(t)\, a(t-\tau)\cos(\omega t + \varphi)\, dt, \tag{13.27}$$

considered as a function of phase φ.

We will show that function $\xi(\tau, \varphi)$ may be obtained by passing oscillation y(t) through an optimum linear filter matched with the signal.

When oscillation y(t) is supplied to the input of a linear system, its output voltage equals

$$u_{\text{вых}}(t) = k_4 \int_{-\infty}^{\infty} y(z)\, \eta(t-z)\, dz,$$

where k_4 -- some constant magnitude, while $\eta(t)$ -- linear system pulse characteristic.

In the case of an optimum filter matched with signal $u_c(t)$,

$$\eta(t) = u_c(t_0 - t).$$

Since, in the case being examined, the signal disappears when $t = \tau + \tau_u$, then we select

$$t_0 = \tau + \tau_u;$$

then

$$\eta(t) = u_c(\tau + \tau_{\text{и}} - t)$$

and

$$u_{\text{вых}}(t) = k_4 \int_{-\infty}^{t} y(z)\, u_c(\tau + \tau_{\text{и}} - t + z)\, dz.$$

Since there is no signal when t $< \tau$ (Figure 13.1), then $u_c(t) = 0$ /219
where t $< \tau$ and $u_c(\tau + \tau_{\text{и}} - t + z) = 0$ where $\tau + \tau_{\text{и}} - t + z < \tau$, i. e., where $z < t - \tau_{\text{и}}$; therefore

$$u_{\text{вых}}(t) = k_4 \int_{t-\tau_{\text{и}}}^{t} y(z)\, u_c(\tau + \tau_{\text{и}} - t + z)\, dz.$$

At the moment $t = \tau + \tau_{\text{и}}$, the result is

$$u_{\text{вых}}(\tau + \tau_{\text{и}}) = k_4 \int_{\tau}^{\tau+\tau_{\text{и}}} y(t)\, u_c(t)\, dt = k_4 \int_{\tau}^{\tau+\tau_{\text{и}}} y(t)\, a(t-\tau) \cos(\omega t + \varphi)\, dt.$$

This expression coincides (precise to a constant factor) with (13.27). Consequently, envelope $U_{m\,\text{вых}}(t)$ of optimum filter output voltage at moment $t = \tau + \tau_{\text{и}}$ also coincides (precise to a constant factor) with the envelope with respect to ϕ of function $\xi(\tau, \phi)$, i. e., with magnitude $M_1(\tau)$ (here and elsewhere it is assumed that the filter output voltage envelope exists, i. e., the pulse signal will comprise at least several rf occupation periods). Thus, desired magnitude $M_1(\tau)$ included in expression (13.26) actually is proportional to the optimum filter output voltage envelope at moment $t = \tau + \tau_{\text{и}}$, i. e., at the moment the signal pulse ceases at filter input (Figure 13.1).

But, it was demonstrated in Chapter 2 that, by the moment the signal at optimum filter input ceases, the voltage at its output attains the maximum. Consequently,

magnitude $M_1(\tau)$ is proportional to the peak value of the pulse voltage at optimum filter $O\Phi$ output.

Thus, the basic operation required to determine distribution $P_y(\tau)$ from formula (13.26) and consisting of computation of function $M_1(\tau)$ may be accomplished by using a linear filter matched with the signal and a linear detector separating this filter's output voltage envelope.

Then, if detector output voltage equals $U_{=}(t)$, then

$$M_1(\tau) = k_5 U_{=}(\tau - \tau_{\text{н}}), \tag{13.28}$$

where k_5 -- constant not depending on τ.

Subsequent operations required for $P_y(\tau)$ computation from formula (13.26) present no real difficulties since magnitude N_o and function $P(\tau)$ are assumed known.

Optimum receiver structure especially is simplfied if one assumes a priori distribution $P(\tau)$ is uniform and, if the maximum inverse probability density approach is used, i. e., if one considers true that value τ at which magnitude $P_y(\tau)$ is maximum. In this case, from (13.26) we have

$$P_y(\tau) = k_5 I_0\left[\frac{2M_1(\tau)}{N_0}\right]. \tag{13.29}$$

where k_5 -- constant determined from a normality condition.

Since $I_o(z)$ is a monotonic function of z, then, when τ changes, maximum /220 inverse probability density $P_y(\tau)$ coincides with the function $M_1(\tau)$ maximum. Therefore, when the maximum inverse probability density criterion is used, that value $\tau_{y\text{м}}$ of parameter τ in which function $M_1(\tau)$ is maximum should be used as the true value. Here, the optimum receiver boils down to optimum linear filter $O\Phi$ matched with the signal and inertia-free envelope detector $Д$ (Figure 13.2). Since, in this case, the requirement is to determine only the moment of the envelope

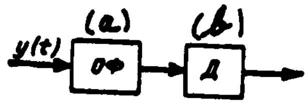


Figure 13.2 (a) -- OF [optimum filter]; (b) -- D [detector].

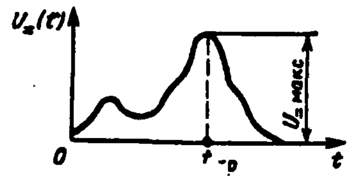


Figure 13.3

maximum, then response curve shape plays no role -- it is only important that this shape be monotonic.

Actually, at some moment t = $t_{\text{кр}}$, let detector output voltage $U_{\text{д}}(t)$ have the greatest value (Figure 13.3). At that moment $t_{\text{кр}}$, the value of envelope voltage $U_{m\,\text{вых}}(t)$ at optimum filter output also will be maximum. Since

$$M_1(\tau) = \text{const}\, U_{m\,\text{вых}}(\tau + \tau_{\text{н}}), \tag{13.30}$$

then function $M_1(\tau)$ will be maximum where $\tau = t_{\text{кр}} - \tau_{\text{н}}$; consequently, the most-probable value $\tau_{y\,\text{н}}$ of the desired parameter is determined from the relationship

$$\tau_{y\,\text{н}} = t_{\text{кр}} - \tau_{\text{н}}. \tag{13.31}$$

(Here, just as before, index y denotes that the magnitude of the most-probable parameter value also will differ for different y(t) realizations).

Since pulse duration $\tau_{\text{н}}$ is assumed known, then it suffices for determination of magnitude $\tau_{y\,\text{н}}$ to determine moment $t_{\text{кр}}$, when receiver output voltage $U_{\text{д}}(t)$ attains the greatest value (Figure 13.3).

Thus, for the given assumptions, measurement of time τ of pulse arrival at receiver input boils down to determination of moment $t_{\text{кр}}$ when receiver output voltage $U_{\text{д}}(t)$ reaches the greatest value.

As already noted above, the maximum inverse probability density criterion is not the best, in particular does not insure the least mean square measurement error. However, it will be demonstrated later (Chapter 19) that, given a high signal-to-noise ratio, i. e., given a requirement for high measurement precision, this criterion is the best. Therefore, we will examine the case of a high /221 signal-to-noise ratio in more detail. Here, in accordance with expression (13.7), it is possible to assume

$$P_y(\tau) = k_5 \frac{e^{2M_1(\tau)/N_0}}{\sqrt{2\pi \frac{2M_1(\tau)}{N_0}}}.$$

Since, given a high signal-to-noise ratio, the indicator of the exponential factor power is much greater than unity, this factor, given τ changes, changes much faster than the expression in the denominator. Therefore, it is possible to assume

$$P_y(\tau) = k_6 e^{2M_1(\tau)/N_0}, \tag{13.32}$$

where constant factor k_6 may be determined from a normality condition.

In accordance with expression (13.27), function $M_1(\tau)$ is the envelope with respect to φ function

$$\xi(\tau, \varphi) = \int_{\tau}^{\tau+\tau_u} y(t)\, a(t-\tau) \cos(\omega t + \varphi)\, dt. \tag{13.27}$$

Here

$$y(t) = a(t-\tau_0) \cos(\omega t + \varphi_0) + u_m(t), \tag{13.33}$$

where τ_0 and φ_0 -- parameter τ and φ true values for a signal present at input during a given observation cycle.

We will assume that

$$\omega\tau_{\text{и}} \gg 1, \tag{13.34}$$

i. e., that the signal pulse will comprise at least several rf occupation periods. Then, when expression (13.33) is substituted into (13.27), it is possible without disrupting the generality of the result to assume that $\varphi_0 = 0$.

In addition, when using formulas (13.27) and (13.33) for a relatively-high signal-to-noise ratio in computation of $\xi(\tau, \varphi)$, it is possible to ignore the second term in expression (13.33). Then, expression (13.27) takes the form

$$\xi(\tau, \varphi) = \int_{\tau}^{\tau+\tau_{\text{и}}} a(t-\tau_0)\, a(t-\tau) \cos \omega t \cos(\omega t + \varphi)\, dt. \tag{13.35}$$

Considering inequality (13.34), it is not difficult to demonstrate that

$$\xi(\tau, \varphi) = \left[\frac{1}{2} \int_{\tau}^{\tau+\tau_{\text{и}}} a(t-\tau_0)\, a(t-\tau)\, dt\right] \cos \varphi$$

and, consequently,

$$M_1(\tau) = \frac{1}{2} \int_{\tau}^{\tau+\tau_{\text{и}}} a(t-\tau_0)\, a(t-\tau)\, dt. \tag{13.36}$$

Since, given a high signal-to-noise ratio, the error in measurement of /222 magnitude τ is slight, curve $P_y(\tau)$ has significant values only in that region where τ approximates τ_0, i. e., given slight values of difference

$$\Delta\tau = \tau - \tau_0; \tag{13.37}$$

therefore, it is possible to assume

$$a(t-\tau) = a(t-\tau_0) + a'(t-\tau_0)\,\Delta\tau + \frac{1}{2}\, a''(t-\tau_0)\,(\Delta\tau)^2, \tag{13.38}$$

where

$$\left.\begin{aligned} a'(t-\tau_0) &= \left(\frac{\partial a(t-\tau)}{\partial \tau}\right)_{\tau=\tau_0}; \\ a''(t-\tau_0) &= \left(\frac{\partial^2 a(t-\tau)}{\partial \tau^2}\right)_{\tau=\tau_0}. \end{aligned}\right\} \tag{13.39}$$

Substituting these expressions into (13.36), we obtain

$$\begin{aligned} M_1(\tau) = Q + \frac{\Delta\tau}{2} \int\limits_{\tau}^{\tau+\tau_H} a(t-\tau_0)\, a'(t-\tau_0)\, dt + \\ + \frac{(\Delta\tau)^2}{2} \int\limits_{\tau}^{\tau+\tau_H} a(t-\tau_0)\, a''(t-\tau_0)\, dt, \end{aligned} \tag{13.40}$$

where

$$Q = \frac{1}{2} \int\limits_{\tau}^{\tau+\tau_H} a^2(t-\tau_0)\, dt$$

is signal energy not depending on τ .

But

$$\int\limits_{\tau}^{\tau+\tau_H} a(t-\tau_0)\, a'(t-\tau_0)\, dt = -\int\limits_{\tau}^{\tau+\tau_H} a(t-\tau_0) \frac{da(t-\tau_0)}{dt}\, dt = \\ = \Big|_{\tau}^{\tau+\tau_H} \frac{a^2(t-\tau_0)}{2} = 0,$$

while

$$\int_{\tau}^{\tau+\tau_{и}} a(t-\tau_0)\,a''(t-\tau_0)\,dt = \int_{\tau}^{\tau+\tau_{и}} a(t-\tau_0)\,\frac{d^2 a(t-\tau_0)}{dt^2}\,dt;$$

performing integration by parts, we obtain

$$\int_{\tau}^{\tau+\tau_{и}} a(t-\tau_0)\,a''(t-\tau_0)\,dt = \Big|_{\tau}^{\tau+\tau_{и}} a(t-\tau_0)\,\frac{da(t-\tau_0)}{dt} - \int_{\tau}^{\tau+\tau_{и}} \left[\frac{da(t-\tau_0)}{dt}\right]^2 dt = 0 - \int_{\tau}^{\tau+\tau_{и}} [a'(t-\tau_0)]^2\,dt.$$

Considering these results, expression (13.40) takes the form /223

$$M_1(\tau) = Q - \frac{(\Delta\tau)^2}{4}\int_{\tau}^{\tau+\tau_{и}} [a'(t-\tau)]^2\,dt;$$

therefore, formula (13.32) may be written in the following manner:

$$P_y(\tau) = k_7\,e^{-B(\tau-\tau_0)^2}, \tag{13.41}$$

where

$$B = \frac{1}{2N_0}\int_{\tau}^{\tau+\tau_{и}} \left[\frac{\partial a(t-\tau)}{\partial\tau}\right]_{\tau_0}^2 dt. \tag{13.42}$$

It follows from the normality condition that

$$\int_0^T P_y(\tau)\,d\tau = 1,$$

i. e.,

$$k_7 \int_0^T e^{-B(\tau-\tau_0)^2}\,d\tau = 1. \tag{13.43}$$

Considering that, given a high signal-to-noise ratio, factor B is large, it is possible in expression (13.43) to replace the finite limits with infinite limits, i. e., to assume

$$k_7 \int_{-\infty}^{\infty} e^{-B(\tau-\tau_0)^2} d\tau = 1;$$

the result here is

$$k_7 = \sqrt{\frac{B}{\pi}}$$

and

$$P_y(\tau) = \sqrt{\frac{B}{\pi}} e^{-B(\tau-\tau_0)^2}. \qquad (13.44)$$

It follows from this formula that distribution $P_y(\tau)$ is a gaussian curve having a maximum when $\tau = \tau_0$, i. e., for the true value of the measured parameter. However, instead of (13.44), it is more correct to assume

$$P_y(\tau) = \sqrt{\frac{B}{\pi}} e^{-B(\tau-\tau_{y\,n})^2}, \qquad (13.45)$$

where $\tau_{y\,n}$ -- most-probable value [i. e., value corresponding to the maximum of curve $P_y(\tau)$], which approximates τ_0, but it is not mandatory that it coincide precisely with it.

Actually, it was pointed out above that the optimum receiver supplies /224 voltage proportional to $P_y(\tau)$; therefore, in principle, it is possible absolutely precisely to determine value $\tau_{y\,n}$, which corresponds to the curve $P_y(\tau)$ maximum. If $\tau_{y\,n}$ always coincides with τ_0, then this would signify that the measurement error always equals zero. But, this contradicts expression (13.44) since it is evident from it that distribution $P_y(\tau)$ differs from the delta-function and, consequently, there always is a finite probability of some measurement error values.

This contradictory result was obtained because, when computing $M_1(\tau)$, we assumed in expression (13.33) a noise voltage not only slight compared to the signal, but also identical with zero. This led to the fact that, instead of approximate value $\tau_{y\pi}$ of the measured parameter, its true value τ_0 was included in expression (13.44).

Thus, if the signal-to-noise ratio is high, but does not equal infinity, distribution $P_y(\tau)$ is determined from formula (13.45) rather than (13.44).

We will compare expression (13.45) obtained for a random initial phase signal with the corresponding expression for a precisely-known signal.

Given a high signal-to-noise ratio and a precisely-known signal, relationships (6.12) are valid, i. e.,

$$\left.\begin{aligned} P_y(x) &= \sqrt{\frac{b}{\pi}}\, e^{-b(x-x_{y\pi})^2}, \\ \text{where}\quad b &= \frac{1}{N_0}\int_0^T \left[\frac{\partial u_x(t)}{\partial x}\right]^2 dt. \end{aligned}\right\} \tag{6.12}$$

Here, x — normalized value of the measured parameter, i. e.,

$$\tau = \tau_0(1 + xm), \tag{13.46}$$

where

$$-1 < x < 1.$$

Replacing variable x with τ in (6.12), in accordance with expressions (8.1), (8.2), and (13.46), we will obtain

$$P_y(\tau) = \sqrt{\frac{b_1}{\pi}}\, e^{-b_1(\tau-\tau_{y\pi})^2},$$

where $$b_1 = \frac{1}{N_0} \int\limits_0^T \left[\frac{\partial u_\tau(t)}{\partial \tau} \right]^2 dt. \tag{13.47}$$

Initially, we will examine that signal $u_\tau(t)$ for which only the envelope will depend on τ, i. e.,

$$u_\tau(t) = a(t-\tau)\cos(\omega t + \varphi), \quad \text{where} \quad \tau \leqslant t \leqslant \tau + \tau_n;$$ /225

(13.48)

$$u_\tau(t) = 0, \text{ outside these limits}$$

Substituting (13.48) into (13.47) and assuming that $\omega T \gg 1$, we will obtain

$$P_y(\tau) = \sqrt{\frac{b_1}{\pi}}\, e^{-b_1(\tau-\tau_{y\,n})^2},$$

where

$$b_1 = \frac{1}{2N_0} \int\limits_\tau^{\tau+\tau_n} \left[\frac{\partial a(t-\tau)}{\partial \tau} \right]^2 dt. \tag{13.49}$$

Comparing this result with corresponding formulas (13.42) and (13.45) obtained above for an identical signal, but which has random phase φ, we will be convinced of their complete coincidence.

Consequently, if measured parameter τ is contained only in the signal envelope and the signal-to-noise ratio is high, parameter measurement accuracy will not depend on whether signal initial phase φ is precisely known or random. Therefore, formulas (6.16) and (6.17) are valid for the mean square of the measurement error and it is possible to assume:

$$\begin{aligned} \overline{\delta^2} &= \overline{(x - x_{y\,n})^2} = \frac{1}{2b}; \\ \overline{(\Delta\tau)^2} &= \overline{(\tau - \tau_{y\,n})^2} = \frac{1}{2b_1} = m^2 \tau_0^2 \overline{\delta^2}. \end{aligned} \tag{13.50}$$

It follows from relationship (13.46) and Figure 13.1 that, when $\tau_{и} \ll T$, $m\tau_1 \approx \frac{T}{2}$ must be the case; here

$$\overline{(\Delta\tau)^2} \approx \frac{T^2}{4}\overline{\delta^2}. \tag{13.50a}$$

If envelope a(t) is described by formula (6.35), then formula (6.40) is valid for $\overline{\delta^2}$, to wit

$$\overline{\delta^2} = \frac{6N_0}{\Omega^2 T^2 Q}. \tag{6.40}$$

We now will examine a signal of the type

$$\left.\begin{aligned} &u_\tau(t) = a(t-\tau)\cos[\omega(t-\tau)+\varphi], \text{ where } \tau \leqslant t \leqslant \tau+\tau_{и}; \\ &u_\tau(t) = 0, \text{ outside these limits} \end{aligned}\right\} \tag{13.51}$$

Measured parameter τ of such a signal modulates both signal amplitude, as well as its phase.

If initial phase φ is random and equally probable in the $0-2\pi$ range, then, instead of (13.51), it is possible to write

$$u_\tau(t) = a(t-\tau)\cos(\omega t+\varphi_1), \text{ where } \tau \leqslant t \quad \tau+\tau_{и}, \tag{13.52}$$

where magnitude $\varphi_1 = \omega\tau + \varphi$ also is random and equally probable /226
in the $0-2\pi$ range.

Consequently, if phase φ is random and equally probable in the $0-2\pi$ range, a type (13.51) signal does not differ at all from the type (13.48) signal examined above. Therefore, it remains to examine a case when a type (13.51) is precisely known (with the exception of τ).

Here, substituting (13.52) into (13.47) and assuming $\omega T \gg 1$, we will obtain

$$\left.\begin{aligned} P_y(\tau) &= \sqrt{\frac{b_1}{\pi}}\, e^{-b_1(\tau-\tau_{y\,\text{n}})^2}, \\ b_1 &= \frac{1}{N_0}\left\{\frac{1}{2}\int_{\tau}^{\tau+\tau_{\text{n}}}\left[\frac{\partial a(t-\tau)}{\partial\tau}\right]^2 dt + \omega^2 Q\right\}. \end{aligned}\right\} \tag{13.53}$$

where

In accordance with formula (13.50), the mean square of the error equals

$$\overline{(\Delta\tau)^2} = \frac{1}{2b_1}. \tag{13.54}$$

It follows from formulas (13.53) and (13.54) that, generally speaking, it is possible by increasing frequency ω to obtain a measurement error as slight as desired. However, realization of such high accuracy may turn out to be very difficult for the following reasons:

1. Initial phase φ must be precisely known at the point of reception, which essentially does not occur in a majority of cases.

2. Increased measurement accuracy obtained by increasing ω is achieved without changing signal energy or broadening its spectrum. But, Kotel'nikov demonstrated (see § 6.3) that, in this case, the minimum signal-to-noise ratio at which the obtained results still remain valid increases. Therefore, the decrease in measurement error due to an increase in frequency ω (i. e., due to decrease in the signal rf occupation period) may be realized only for very high input signal levels or for a very low noise level.

For these reasons, in a majority of essentially interesting cases, minimum permissible τ measurement error is determined from formulas (13.49) and (13.50), namely:

$$\overline{(\Delta\tau)^2} = \frac{N_0}{\int_{\tau}^{\tau+\tau_{\text{n}}}\left[\frac{\partial a(t-\tau)}{\partial\tau}\right]^2 dt}. \tag{13.55}$$

If the pulse envelope is described by formula (6.35), i. e., the pulse has a rectangular spectrum, bounded by band $\omega = (\omega_0 - \Omega) - (\omega_0 + \Omega)$, then, in accordance with (13.50a) and (6.40), the result is

$$\overline{(\Delta\tau)^2} = \frac{T^2}{4}\overline{\delta^2} = \frac{3N_0}{2Q\Omega^2}. \tag{13.56}$$

Formulas (13.55) and (13.56) are equally valid for precisely-known /227 and for random initial phase signals when the signal-to-noise ratio is high.

Sometimes, it is more convenient to write formula (13.55) in the following form:

$$\overline{(\Delta\tau)^2} = \frac{N_0}{\frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \omega^2 |G_n(j\omega)|^2 d\omega}, \tag{13.57}$$

where $G_n(j\omega)$ -- envelope a(t) complex spectrum (Fourier transform).

Since signal energy equals

$$Q = \frac{1}{2}\int\limits_{-\infty}^{\infty} a^2(t)\,dt = \frac{1}{4\pi}\int\limits_{-\infty}^{\infty} |G_n(j\omega)|^2 d\omega, \tag{13.58}$$

then it is possible to write formula (13.57) in the following form:

$$\overline{(\Delta\tau)^2} = \frac{N_0}{2Q\beta^2}, \tag{13.59}$$

where

$$\beta^2 = \frac{\int\limits_{-\infty}^{\infty} \omega^2 |G_n(j\omega)|^2 d\omega}{\int\limits_{-\infty}^{\infty} |G_n(j\omega)|^2 d\omega}. \tag{13.60}$$

The result for the signal examined above with a rectangular envelope frequency spectrum in the $-\Omega$ to Ω range is

$$\beta^2 = \frac{1}{3}\Omega^2, \tag{13.61}$$

and formula (13.59) is transformed into formula (13.56).

13.3 Measurement of Random-Phase Signal Frequency

In observation cycle (0, T), let the signal have the form

$$u_c(t) = a(t)\cos(\omega t + \varphi), \tag{13.62}$$

where a(t) -- known time function; ω -- measured parameter with a priori distribution P(ω); φ -- parasitic random parameter with distribution (9.3).

Accomplishing insertions analogous to those made in the preceding section, instead of (13.25)--(13.27), we will obtain the following relationships:

$$P_y(\omega) = k_1 P(\omega) I_0\left[\frac{2M_2(\omega)}{N_0}\right], \tag{13.63}$$

where k_1 -- normalizing constant;

/228

$$\left.\begin{aligned} M_2(\omega) &= \sqrt{X_2^2 + Y_2^2}; \\ X_2 &= \int_0^T y(t)\, a(t)\cos\omega t\, dt; \\ Y_2 &= \int_0^T y(t)\, a(t)\sin\omega t\, dt, \end{aligned}\right\} \tag{13.64}$$

or $M_2(\omega)$ is the envelope with respect to φ of oscillation

$$\xi_1(\omega, \varphi) = \int_0^T y(t)\, a(t)\cos(\omega t + \varphi)\, dt. \tag{13.65}$$

It follows from (13.65) that magnitude $M_2(\omega)$ for every value of frequency ω is proportional to the value of the envelope (at moment t = T) of the oscillation at output of an optimum linear filter matched with signal

$$u_c(t) = a(t)\cos(\omega t + \varphi).$$

Absence of data on signal phase φ is no obstacle to design of this filter since the filter at output must supply the correct value only of the rf oscillation envelope. Therefore, it is possible to assume any phase φ, such as one equalling zero, during filter design.

The main difficulty is that function $P_y(\omega)$ and, consequently, function $M_2(\omega)$ as well, must be computed not only for one ω value, but for an infinite multitude of ω values ranging from $2\pi f_{мин}$ to $2\pi f_{макс}$, where $f_{мин}$ and $f_{макс}$ -- limits of possible frequency f changes in the anticipated signal ($f = \omega/2\pi$).

Since optimum filter structure significantly will depend on signal carrier frequency f, then, strictly speaking, a set of an infinite number of such filters is required. (If realization y(t) may be written and retained for a long time, then it is possible to use one filter with tunable center frequency f of this filter's bandwidth.)

No similar difficulty arises when measuring signal amplitude or its lag τ since optimum linear filter structure will not depend on signal amplitude a or its time of arrival τ. The aforementioned difficulty occurs in a case of frequency measurement and, to overcome this, anticipated signal pass-band $f_{мин} - f_{макс}$ must be divided into n sections of finite width

$$\Delta f = \frac{f_{макс} - f_{мин}}{n},$$

where Δf -- magnitude less than the width of the basic portion of the anticipated signal spectrum. Here, instead of an infinite number of filters, it suffices

to have a finite number n of such filters, which makes it possible to /229
compute a sufficient number of points $P_y(\omega_1)$, $P_y(\omega_2)$, . . ., $P_y(\omega_n)$ to plot desired curve $P_y(\omega)$.

We now will examine the problem of ultimate frequency measurement precision when the signal-to-noise ratio is high. Here, as in previous cases, we will assume that a priori distribution $P(\omega)$ of the measured parameter is uniform. Then, analogous to (13.32) we obtain

$$P_y(\omega) = k_2 e^{\frac{2M_2(\omega)}{N_0}}, \qquad (13.66)$$

where k_2 -- normalizing constant, while $M_2(\omega)$ is the envelope with respect to φ of function $\xi_1(\omega, \varphi)$, determined from expression (13.65).

It is possible, given a high signal-to-noise ratio, to assume in expression (13.65) in the first approximation

$$y(t) = a(t)\cos(\omega_0 t + \varphi_0),$$

where ω_0 and φ_0 -- true values of signal frequency and initial phase occurring during a given observation cycle (0, T).

Then, from (13.65), we will obtain

$$\xi_1(\omega, \varphi) = \int_0^T a^2(t)\cos(\omega_0 t + \varphi_0)\cos(\omega t + \varphi)\,dt.$$

Evidently, without disrupting the generality, it is possible to assume $\varphi_0 = 0$ in this expression; then

$$\xi_1(\omega, \varphi) = X\cos\varphi - Y\sin\varphi$$

and

$$M_2(\omega) = \sqrt{X^2 + Y^2}, \tag{13.67}$$

where

$$\left.\begin{aligned} X &= \int_0^T a^2(t) \cos \omega_0 t \cos \omega t \, dt; \\ Y &= \int_0^T a^2(t) \cos \omega_0 t \sin \omega t \, dt. \end{aligned}\right\} \tag{13.68}$$

We will designate

$$\Delta\omega = \omega - \omega_0 \tag{13.69}$$

and we will assume that

$$\left.\begin{aligned} &(\omega_{\text{макс}} - \omega_{\text{мин}}) T < 1, \\ &\text{and, consequently, and} \\ &\Delta\omega T < 1, \end{aligned}\right\} \tag{13.70a}$$

and that

$$\omega_0 T \gg 1. \tag{13.70b}$$

It follows from (13.70a) that, in formulas (13.68), it is possible to assume /230

$$\left.\begin{aligned} \cos \Delta\omega t &\approx 1 - \frac{(\Delta\omega t)^2}{2}, \\ \sin \Delta\omega t &\approx \Delta\omega t. \end{aligned}\right\} \tag{13.71}$$

Considering relationships (13.69), (13.70b), and (13.71), it is possible

to use relatively-unwieldly but uncomplicated transformations to reduce formulas (13.67) and (13.68) to the following form:

$$M_3(\omega)=Q-\frac{b_2}{2}(\omega-\omega_0)^2, \tag{13.72}$$

where

$$Q=\frac{1}{2}\int_0^T a^2(t)\,dt \tag{13.73}$$

is relative signal energy, while

$$b_2=\frac{1}{2}\int_0^T a^2(t)\,t^2\,dt-\frac{1}{4Q}\left[\int_0^T a^2(t)\,t\,dt\right]^2. \tag{13.74}$$

Substituting expression (13.72) into (13.66), we will obtain

$$P_y(\omega)=k_3 e^{-\frac{b_2}{N_0}(\omega-\omega_0)^2}, \tag{13.75}$$

where k_3 -- normalizing constant.

For the identical reasons which led to replacement of expression (13.44) by (13.45), it is more correct to replace true signal frequency value ω_0 in (13.75) by its most-probable value $\omega_{y\,н}$, i. e., to assume

$$P_y(\omega)=k_3 e^{-\frac{b_2}{N_0}(\omega-\omega_{y\,н})^2}.$$

Considering the normality condition, i. e., assuming

$$k_3\int_{\omega_{мин}}^{\omega_{макс}} e^{-\frac{b_2}{N_0}(\omega-\omega_{y\,н})^2}\,d\omega\approx k_3\int_{-\infty}^{\infty} e^{-\frac{b_2}{N_0}(\omega-\omega_{y\,н})^2}\,d\omega=1,$$

we will obtain

$$k_2 = \sqrt{\frac{N_0}{b_2 \pi}}$$

and

$$P_y(\omega) = \sqrt{\frac{N_0}{b_2 \pi}} \, e^{-\frac{b_2}{N_0}(\omega - \omega_{y\,n})^2} \tag{13.76}$$

Consequently, given a high signal-to-noise ratio, distribution $P_y(\omega)$, just like $P_y(a)$ and $P_y(\tau)$ are described by a gaussian curve. Therefore, the mean square of the frequency measurement error equals

$$\overline{(\Delta\omega)^2} = \frac{N_0}{2b_2}, \tag{13.77}$$

where b_2 is determined by formulas (13.73) and (13.74). /231

It was noted in Chapter 6 in the examination of the problem of measuring the frequency of a precisely-known signal that measurement error will depend on moment t_1, at which observation begins. Therefore, in order to clarify whether or not an analogous phenomenon exists in the case under examination, we must in expressions (13.73) and (13.74) assume the observation cycle equals $t_1 - t_1 + T$, rather than (0, T). Then, we will obtain

$$b_2 = \frac{1}{2} \int_{t_1}^{t_1+T} a^2(t)\, t^2\, dt - \frac{1}{4Q} \left[\int_{t_1}^{t_1+T} a^2(t)\, t\, dt \right]^2, \tag{13.78}$$

where

$$Q = \frac{1}{2} \int_{t_1}^{t_1+T} a^2(t)\, dt.$$

Let signal amplitude in the observation cycle be constant, i. e., $a(t) = a_0$; then, from (13.77) and (13.78), we will obtain

$$\overline{(\Delta\omega)^2} = \frac{12N_0}{a_0^2T^3} = \frac{6N_0}{QT^2}. \tag{13.79}$$

If normalized message x is used, as was the case in Chapter 6, then, instead of (13.79), we will obtain

$$\overline{\delta^2} = \frac{\overline{(\Delta\omega)^2}}{\Omega^2} = \frac{6N_0}{Q\Omega^2T^2}. \tag{13.80}$$

It follows from formulas (13.79) and (13.80) that, given a signal with random equiprobable (ranging from $0 - 2\pi$) initial phase, the frequency measurement error will not depend on moment t_1 when observation begins.

Comparison of formulas (13.80) and 6.28b) demonstrates that resultant error during measurement of a random-phase signal frequency is identical to that obtained in the worst case (i. e., where $t_1 = T/2$) for a precisely-known signal.

This result is understandable since, in Chapter 6, it was indicated that, in the case of a precisely-known signal, a decrease in measurement error obtained when t_1 is increased may be realized only if the initial phase of signal rf occupation is precisely known at the point of reception.

13.4 Simultaneous Signal Detection and Measurement of Its Parameters

It was assumed in previous sections of this chapter that realization y(t), the basis for analysis of which signal parameters (amplitude, frequency, and so forth) must be reproduced, must contain the signal. But, in radar and in certain other instances, there often is a requirement to use analysis of this same realization y(t) to solve the problem of both signal parameter measurement and its /232 detection when it is not known beforehand whether or not a given realization contains signal-plus-noise or just noise.

This problem of simultaneous signal detection and measurement of its analog

parameters to a significant degree is similar to the problem of simultaneous detection and recognition of m equiprobable orthogonal signals examined in Chapter 11.

Actually, let the problem be, for instance, simultaneous, i. e., with respect to the same realization y(t), determination whether or not a given realization contains a signal pulse and, if so, what this signal's lag τ (moment of arrival) is. The measured parameter's a priori distribution equals P(τ) and it is known that τ ranges from $\tau_{мин} - \tau_{макс}$. The anticipated signal's duration is $\tau_и$.

Parameter τ is an analog magnitude but, in the first approximation, it is possible to assume that τ may accept only discrete values with interval $\Delta\tau = \tau_и$.

Number m of such discrete values equals

$$m = \frac{T}{\tau_и}, \tag{13.81}$$

where T — observation cycle, which must encompass all possible signal pulse positions. Therefore

$$T = (\tau_{макс} - \tau_{мин}) + \tau_и$$

and

$$m = 1 + \frac{\tau_{макс} - \tau_{мин}}{\tau_и}. \tag{13.82}$$

Since, given such assumptions, possible pulse signals do not overlap over time, then they are orthogonal and the parameter τ measurement problem will boil down to recognition of which of m possible orthogonal signals will be contained in given realization y(t). Consequently, the aforementioned parameter digitization reduces the problem of simultaneous signal detection and measurement of its parameter τ to a problem of simultaneous detection and recognition of m orthogonal signals.

Parameter τ in the general case may have irregular a priori distribution $P(\tau)$. But, as was mentioned often above and will be demonstrated in Chapter 19, in the majority of real cases and, in particular, given high parameter measurement accuracy, it is possible to assume that distribution $P(\tau)$ is uniform. Here, evidently, it is possible accordingly to assume that all m orthogonal signal values are equally probable.

Thus, the task of simultaneous signal parameter detection and measurement boils down in the first approximation to the problem of simultaneous detection and recognition of m equally-probable orthogonal signals. This problem was examined in detail in Chapters 11 and 12, with the following basic results obtained:

1. In the general case, optimum signal detection without its recognition /233 requires less signal energy and another optimum system structure than does simultaneous signal detection and recognition. Therefore, in the general case, if analysis of given realization y(t) is the foundation for obtaining the most-reliable results in the case of both detection and recognition, there must be two receiving systems, an optimum detection (without recognition) and an optimum recognition (without detection) system. Here, the results of the second system are considered only in those cases when the first system establishes, with sufficient reliability, that a signal is present (some elements of both systems may be common, but output elements must be different).

However, when there are high requirements for detection reliability (there can be no highly-reliable recognition without highly-reliable detection), simultaneous detection and recognition requires only slightly-greater signal energy than is the case for detection without recognition. Therefore, when detection and recognition reliability requirements are high, it is inadvisible to create a separate system for optimum detection (without recognition).

2. A change in the number m of orthogonal signals by a factor of 10—100 has virtually no impact on requisite signal energy if the permissible false-alarm probability is slight ($P_{лт} \leqslant 10^{-3}$) or if m $\geqslant 10^4$ [see formulas (11.27)--(11.29)].

3. If $P_{лт} \leqslant 0{,}1$ (which usually is the case), energy required for detection

with recognition of m equally-probable orthogonal signals may be determined from simple binary detection formulas, if $P_{лт}$ in them is replaced by $P_{лт}/m$.

4. If $P_{лт} \leqslant 0{,}1$ and $P_{лт} \leqslant 0{,}1 P_{пр}$ (which usually is the case), then the result is

$$P_{иск} \approx P_{пр},$$

i. e., recognition error probability (given that some sort of signal is present) equals miss probability.

Considering the above analogy between recognition of orthogonal signals and measurement of analog parameter τ , it is possible to formulate the following postulations valid for simultaneous signal detection and measurement of its analog parameter τ :

1'. In the general case, two optimum systems are needed for given realization y(t) for optimum solution of the problem of signal detection and measurement of its parameter.

The first must provide optimum solution of the detection (without recognition) problem, while the second accomplishes optimum parameter measurement (without detection). Here, second system data are considered only in those cases when the first system establishes with sufficient reliability that a signal is present. (Some units in both systems are common, but output units must be separate).

However, given high requirements for detection reliability (but, there /234 may be no highly-reliable parameter measurement without highly-reliable detection), simultaneous parameter detection and measurement provide the same detection reliability and require only slightly-more signal energy than is the case for parameter detection without measurement.

Therefore, it is inadvisible to create a separate system for optimum detection (without parameter measurement), given high requirements for parameter detection and measurement reliability.

2'. Given high requirements for parameter measurement accuracy, the optimum system is the one founded on the principle of greatest inverse probability density $P_y(\tau)$, i. e., the system supplying that parameter value $\tau = \tau_{y\,н}$, for which density $P_y(\tau)$ is maximum.

As noted earlier and is demonstrated in Chapter 19, this system is optimum in the sense that it provides the minimum mean square measurement error and in the sense of a broad class of other optimizations.

It follows from the above point that it is inadvisible to create a separate system for optimum detection (without parameter measurement), given high requirements for parameter detection and measurement reliability. Therefore, an optimum simultaneous parameter detection and measurement system must operate in the following manner (similar to the action of the Figure 11.2 optimum system for detection with recognition of m orthogonal signals). It must determine the greatest value $P_y(\tau_{y\,н})$ of the measured parameter inverse probability density and compare it with some threshold U_0 selected in accordance with permissible false-alarm probability $P_{лт}$. If it turns out that

$$P_y(\tau_{y\,н}) \leqslant U_0,$$

then the decision is that no signal is present; if

$$P_y(\tau_{y\,н}) > U_0,$$

then the decision is that a signal is present at input and that $\tau_{y\,н}$ is the desired value of measured parameter τ. Thus, for example, the system for optimum measurement depicted in Figure 13.2 will supply optimum signal detection and measurement, if one assumes that a signal is present only in those cases when maximum detector output voltage value $U_{m\,макс}$ (Figure 13.3) exceeds some predetermined threshold U_0. The methodology for measurement of τ remains as before, i. e., it is assumed:

$$\tau = t_{нр} - \tau_н.$$

3'. It is permissible to replace a real signal with analog parameter τ of a system of m equally-probable orthogonal signals to determine the signal energy required for reliable signal detection in a system of simultaneous signal /235 detection and measurement of its parameter. Magnitude m may be determined approximately from formula (13.82).

The circumstance that magnitude m may be determined only approximately (since m orthogonal signals only are approximately equivalent to one signal with analog parameter τ), does not play a large role since, as noted above (in point 2), in a majority of cases, an error in determination of m by a factor of 10 or even of 100 is permissible.

4'. If $P_{лт} \leq 0.1$ (which usually is the case), then the energy required for signal detection with given error probabilities $P_{лт}$ and $P_{пр}$ may be determined from simple binary detection formulas if $P_{лт}$ in them is replaced by $P_{лт}/m$.

5'. It follows from the above points that, during simultaneous signal detection and measurement of its parameter, the energy required to provide given detection error probabilities $P_{лт}$ and $P_{пр}$ (given slight values for these probabilities) may be computed from formulas presented in previous chapters for a case of simultaneous detection and recognition of m equiprobable orthogonal signals.

Energy required to provide the given parameter measurement mean square error is determined from formulas presented in this chapter [from formula (6.40), for example].

In individual cases, requisite energy will be determined in the final analysis with permissible detection error probability values ($P_{лт}$ and $P_{пр}$), while, in other cases, with the permissible parameter measurement error.

All the above discussions for simplicity were applied to a case of measuring lag τ . But, it is evident that they may be used to measure other signal parameters as well. For example, if the requirement is to carry out simultaneous detection of a signal and measurement of its frequency, then the only difference will be that number m of the equivalent orthogonal signals must be determined, not from formula (13.82), but from the formula

$$m = 1 + \frac{f_{макс} - f_{мин}}{\Delta f}, \qquad (13.83)$$

where Δf -- width of the basic part of the signal spectrum (i. e., of that part of the spectrum in which the main part of the signal energy is concentrated, for example, 70—80% of this energy), while $f_{мин}$ and $f_{макс}$ -- of possible measured frequency f change.

We will examine as our example radar detection and recognition of targets located at different azimuths and ranges. Azimuth θ and lag τ (proportional to range r) of each target are analog parameters, which may change in the range $\Theta_{мин} - \Theta_{макс}$ and $\tau_{мин} - \tau_{макс}$. However, when determining the energy required for detection of a target with given error probabilities $P_{лт}$ and $P_{пр}$, we will assume that the signal reflected from each target is equivalent to a signal which may have one of m equally-probable orthogonal values, where

/236

$$\left.\begin{aligned} m &= m_1 m_2; \\ m_1 &\approx 1 + \frac{\Theta_{макс} - \Theta_{мин}}{\Delta\Theta}; \\ m_2 &\approx 1 + \frac{\tau_{макс} - \tau_{мин}}{\tau_и}. \end{aligned}\right\} \qquad (13.84)$$

Here, $\Delta\theta$ -- width of the radiation pattern with respect to azimuth, while $\tau_и$ -- pulse signal duration.

Actually, for each position, the radiation patterns of signal pulses shifted relative to one another at intervals $\tau_и, 2\tau_и$ and so on do not overlap over time. Pulses corresponding to radiation pattern positions shifted relative to one another at angle $\Delta\theta$ also do not overlap over time. Therefore, none of the m possible signal puls s, where m is determined from formula (13.84), overlap over time and, consequently, are mutually orthogonal.

Formula (13.84) applies to a case when the frequency of the reflected signal

is precisely known, which usually is the case only for a stationary target. If the target moves and the frequency of the reflected pulse due to the Doppler effect may be found equally probably in the range $_{мин} - f_{макс}$, then, instead of (13.84), it should be assumed that

$$m = m_1 m_2 m_3, \tag{13.85a}$$

where m_1 and m_2, as usual, are determined from formula (13.84), while

$$m_3 \approx 1 + \frac{f_{макс} - f_{мин}}{\Delta f}, \tag{13.85b}$$

where Δf -- width of the basic part of the pulse spectrum.

Δf and τ are linked by the following known relationship for a pulse with sinusoidal rf occupation:

$$\Delta f \approx \frac{1-2}{\tau_и}. \tag{13.85c}$$

It follows from formulas (13.84) and (13.85) that, for short-duration pulses, the result is $m_2 \gg 1$, while $m_3 \approx 1$, i. e., signal frequency irregularity does not increase significantly magnitude m and, consequently, there is no need for a significant increase in the signal energy required for reliable detection. Given long-duration pulses, on the other hand, $m_3 \gg 1$ and the irregularity of the frequency of the signal reflected from the target significantly increases m.

For example, let $\tau_{макс} - \tau_{мин} = 2$ ms; $\tau_и = 2$ us; $\Theta_{макс} - \Theta_{мин} = 360°$; $\Delta\theta = 1°$; $f_{макс} - f_{мин} = 20$ kHz; $\Delta f = 1$ MHz. Then, in accordance with formulas (13.84) and (13.85), we will obtain

$$m_1 \approx 360, \quad m_2 \approx 1\,000, \quad m_3 \approx 1$$

$$m \approx m_1 m_2 \cdot 1 \approx 3{,}6 \cdot 10^5.$$

If signal amplitude is assumed to be known, then signal energy required /237 for detection with error probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$ may be determined from formula (11.28), i. e.,

$$\frac{Q}{N_0}=\left(\sqrt{\ln\frac{1}{P_{\text{лт}}}+\ln m}+\sqrt{\ln\frac{1}{P_{\text{пр}}}-1.4}\right)^2. \tag{11.28}$$

This formula applies to a case when the reflected signal comprises one pulse or n coherent pulses.

If the signal takes the form of a packet of n incoherent pulses, then the loss in requisite energy resulting from this incoherence may be determined from formula (12.60) or from the Figure 12.5 curves.

Let $P_{\text{лт}} = 10^{-5}$ and $P_{\text{пр}} = 10^{-1}$. Then, from formula (11.28), we will find

$$\frac{Q}{N_0}=34.$$

We will explain how imprecision in determination of the number m of equivalent orthogonal signals may impact upon ratio Q/N_0 magnitude. We will assume (with a large allowance) that an error by a factor of 100 on either side of the true value may be permitted, i. e., in actuality, m may equal 3.6×10^3 or 3.6×10^7. Then, from formula (11.28), we will obtain, respectively

$$\frac{Q}{N_0}=29, \text{ or } \frac{Q}{N_0}=40,$$

i. e., in this example, an error by a factor of 100 in either direction in the determination of m will lead to an error of less than 0.7 dB in determination of threshold ratio Q/N_0.

In the case of a fluctuating signal, formula (11.29) should be used instead of (11.28), i. e., one should assume

$$\frac{Q_{cp}}{N_0}=\frac{1}{P_{np}}\left(\ln m+\ln\frac{1}{P_{л т}}\right). \qquad (11.29)$$

Where $P_{np}=0,1$ $P_{л т}=10^{-5}$ and m = 3.6 x 10^5, the result is

$$\frac{Q_{cp}}{N_0}=240.$$

In this case, a change (decrease) in m by a factor of 100 also causes only a relatively-slight change Q_{cp}/N_0 of less than 0.9 dB in threshold ratio.

Assuming that signal amplitude and frequency are precisely known, if ratio Q/N_0 is sufficiently high, the parameter τ measurement error may be determined from formula (13.56).

PART FOUR /238

MATHEMATICAL METHODS OF STATISTICS FOR INVESTIGATION OF OPTIMUM SIGNAL RECEPTION

CHAPTER FOURTEEN

SIGNAL DETECTION AND RECOGNITION ANALYSIS USING STATISTICAL HYPOTHESIS TESTING APPROACHES

14.1 General Comments

In preceding chapters, optimum receiver analysis mainly used the maximum inverse probability approach. However, a broad mechanism developed in contemporary mathematical statistics makes it possible significantly to expand and intensify the theory of optimum reception methods and, in particular, to evaluate the applicability range of the maximum inverse probability approach. At present, mathematical statistics is one of the most-rapidly developing branches of mathematics and the results it supplies will find ever-expanding utility in the theory of optimum reception methods.

The main task in mathematical statistics is decision-making based on results of observing random magnitudes (or random functions). It is assumed that the result of observing some random magnitude (or random function) is a system (y_1, y_2, . . ., y_n) of the values of this magnitude (or function).

Each observed value y_i (Figure 14.1) is called a sample value, while the entire system $(y_1, \ldots, y_n)$ is called a sample. The number n of the sample

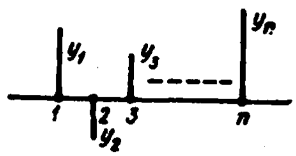


Figure 14.1

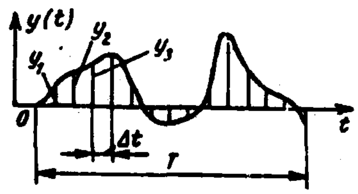


Figure 14.7

values contained in a given sample is called the sample size.

Based on a resultant sample, let the requirement be to establish the law of distribution of the observed random magnitude (or random function). It is evident that the problem may be solved precisely only when $n \longrightarrow \infty$ and, the smaller the n, then, all other conditions being equal, the less the accuracy in representation of the desired law of distribution.

In many cases, the functional shape of the desired law of distribution is known (for example, it is known that this law is normal) and only one or several parameters determining this distribution are unknown (for instance, mean value and dispersion). Then, the problem boils down only to establishing these distribution parameters.

Distribution parameters may be discrete or analog magnitudes. If they /239 are discrete magnitudes, i. e. only discrete values $a_1, a_2, \ldots, a_m$ may be accepted, then, considering that, given a finite sample size, it is impossible to establish with full precision which of these values actually occurs. Only a hypothesis may be formed concerning this or that parameter value. Here, some a posteriori probability of its validity (correctness) corresponds to each such hypothesis, which is called a statistical hypothesis. Consequently, establishing the value of a desired parameter requires solution ("testing") of which of m possible hypotheses ($a = a_1$, $a = a_2$, or $a = a_m$) is most likely for a given sample. Thus, in a case of discrete parameters, the task boils down to testing statistical hypotheses.

If desired parameter a is an analog magnitude, then the task is to use the resultant sample ($y_1, \ldots, y_n$) to evaluate the magnitude of this parameter in some way. The result of this evaluation (in this case a^*) differs from the parameter's true value due to sample size finitude, i. e., there is a parameter measurement error:

$$\Delta a = a^* - a. \tag{14.1}$$

Thus, depending on whether the desired parameter is discrete or analog, the task boils down to testing statistical hypotheses or evaluating the distribution parameter (or several parameters).

It is not difficult to become convinced that these tasks are equivalent to the problems of detection and reproduction of a signal on a noise background. Actually, detection or reproduction of a signal on a noise background requires observation of signal-plus-noise (or of noise alone) realization y(t) to detect whether or not there is a signal and (or) establishing the values of certain signal parameters (amplitude, delay time, and so forth).

In accordance with the Kotel'nikov expansion theorem (see § 1.3), realization y(t) approximately* is characterized by its n values ($y_1, \ldots, y_n$) taken /240 at intervals $\Delta t = 1/2f_n$ (Figure 14.2). Consequently, the observation result may be sample ($y, \ldots, y_n$) with size

$$n = \frac{T}{\Delta t} = 2f_n T. \tag{14.2}$$

The signal-plus-noise realization law of distribution will depend on the signal included within it and, consequently, on signal parameters.

Therefore, if noise parameters are known, then the only unknowns are the signal parameters. Here, finding (determination of) distribution parameters fully equates to determination of signal parameters, i. e., solution of the signal detection or reproduction problem. Only the following comments are required.

*The error of such an approximation is less, the smaller sample size n.

If sample size n strives towards infinity because observation time T rises without restraint [see formula (14.2)], then the error in detection or reproduction problem solution will strive towards zero (if the measured parameter over time T remains unchanged).

If magnitude n strives towards infinity because $\Delta t \longrightarrow 0$ where f_s = const and T = const, then problem solution error not only will not strive towards zero, but remains virtually unchanged (if value n already was sufficiently high prior to being increased). This is because, in accordance with the Kotel'nikov expansion theorem, function y(t) essentially <u>completely</u> is characterized by its n discrete values if n is sufficiently high and spectral energy outside the frequency f_s range is infinitesimally small. This denotes that a decrease in interval Δt among sample values (Figure 14.2) from $1/2f_s$ to zero essentially provides no new information at all about realization y(t) and, consequently, may not increase assigned problem solution accuracy.

If, on the other hand, interval Δt is increased in comparison with $1/2f_s$, then, where T = const, some of the information on function y(t) will be lost and problem solution accuracy decreases.

A sample where $n \longrightarrow \infty$ and T = const $\neq \infty$ (i. e., $\Delta t \longrightarrow 0$ where T = const) is analog rather than discrete, which occurs when $\Delta t \neq 0$.

It follows from what has been stated that an analog sample may not provide significantly-greater problem-solving accuracy than discrete sampling does if, for the discrete sample

$$\Delta t \leqslant \frac{1}{2f_s}. \tag{14.3}$$

Therefore, when condition (14.3) is met, it is immaterial in principle whether a discrete or analog sample is used. Here, sample type is selected only from the point of view of simplicity in mathematical analysis or design accomplishment.

It is evident also that, if some sort of formula is obtained in the /241

form of a sum for the discrete sample, then the corresponding result for an analog sample may be obtained if, in this formula, one assumes:

$$\left.\begin{array}{l} \Delta t \to 0, \\ \text{or} \\ n \to \infty \quad \text{where} \quad T = \text{const.} \end{array}\right\} \qquad (14.4)$$

Here, the sum is transformed into an integral.

Considering these comments, finding distribution parameters for given sample ($y_1, \ldots, y_n$) is equivalent to detection or reproduction of a signal (or of its parameters) for given realization y(t) [at interval (0, T)].

The task for signal detection and recognition is to differentiate among discrete signal types. Therefore, this problem equates to testing statis'ical hypotheses. Measurement of analog signal parameters is equivalent to estimating distribution parameters.

Use of the statistical hypothesis testing approach for signal detection and recognition problems is examined in this and subsequent chapters (Chapters 14 and 15). Chapter 16 is devoted to use of the distribution parameter estimate approach to analog signal parameter reproduction problems.

A more-general theory of statistical solutions is examined in Chapters 17 and 18. This theory, first, makes it possible to approach statistical hypothesis testing and distribution parameter estimation from a common and more-general point of view. Second, it provides the opportunity to solve additional, more complex, statistical problems.

14.2 Signal Detection and Recognition to Test Statistical Hypotheses

As noted in § 8.1, various types of detection are possible.

Initially, we will examine simple binary detection of a precisely-known signal.

Let the amplitude of this signal equal a_o. Then, the detection problem boils down to establishing, using observation of realization y(t) [during interval (0, T)] as the basis, whether this realization contains a signal or corresponds only to noise or, which is the same thing, to establish which of two possible signal amplitude values occurred: a= 0 or $a = a_o$.

This denotes that hypothesis H must be tested (no signal, a = 0) relative to alternative (incompatible) hypothesis H_1 (signal, $a = a_o$).

A signal has not one, but several, possible non-zero values $u_1, u_2, \ldots, u_m$ in complex binary detection and the detection problem boils down to testing hypothesis H_o (no signal of any kind) relative to alternative hypothesis H_1 (one of m possible signals is present). But, here, as opposed to simple binary /242 detection, hypothesis H_1 is called complex.

In detection with recognition of m non-zero signals $u_1, u_2, \ldots, u_m$, m + 1 hypotheses will be subject to testing--H_o (no signal), H_1 (signal u_1 is present), H_2 (signal u_2 is present), . . ., H_m (signal u_m is present). In this case, the zero hypothesis (H_o) is in contrast to m alternative non-zero hypotheses (H_1, . . ., H_m). Therefore, this often is called m-alternative detection.

It was demonstrated in Part III of this book that complex binary detection (i. e., detection without recognition) usually is of less interest and less advisible than simultaneous signal detection and recognition. On the other hand, it was shown that, in the majority of practical cases, energy required for simultaneous detection and recognition of m possible signals may be found from simple binary detection formulas if $P_{л\,т}$ is replaced by $P_{л\,т}/m$ in them. Therefore, for brevity, only simple binary detection is examined below.

14.3 Binary Detection

Initially, we will assume that the signal is precisely known. In this case, the oscillation at receiver input has the form

$$y(t) = u_s(t) + u_{ш}(t). \qquad (14.5)$$

When there is no signal

$$u_x(t) = u_{x_0}(t) \equiv 0;$$

when there is a signal

$$u_x(t) = u_{x_1}(t) = u_c(t), \tag{14.5a}$$

i. e., from the point of view of message reproduction, in this case message x may have only two values:

$$x = x_0 = 0 \text{ and } x = x_1,$$

with a priori probabilities

$$P(x_0) = P(0) \quad \text{and} \quad P(x_1) = P(C)$$

respectively.

One of two mutually-exclusive responses must be provided as a result of observing realization y(t) [during time interval (0, T)]: "yes" (signal, $x = x_1$) or "no" (no signal, $x = x_0$).

Each possible realization $y_i(t)$ represents one point in multidimensional space (see § 6.3). Therefore, from a geometrical point of view, the solution to the

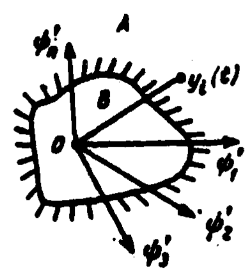


Figure 14.3

detection problem requires that multidimensional space be divided into adjacent, but non-overlapping, regions A and B (Figure 14.3). If "point" $y_i(t)$ falls in region A, the decision is "yes" and the decision is "no" if it does not. /243

Two types of errors are possible when solving this problem--false alarms (with probability $P_{\text{лт}}$) and signal misses (with probability $P_{\text{пр}}$).

Evidently, false alarms occur if realization y(t) falls in region A when there is no signal.

Therefore*

$$P_{\text{лт}} = \int_A P_{x_0}(y)\,dy. \tag{14.6}$$

Signal misses occur if realization y(t) falls in region B when there is a signal; therefore

$$P_{\text{пр}} = \int_B P_{x_1}(y)\,dy. \tag{14.7}$$

Correct detection results if realization y(t) falls in region A when there is a signal. Consequently,

$$P_{\text{по}} = 1 - P_{\text{пр}} = \int_A P_{x_1}(y)\,dy. \tag{14.8}$$

It follows from these formulas or directly from Figure 14.3 that, if region A equals zero, then the result will be:

---

*Here and below, $P_x(y)$ are multidimensional distributions, while $\int_A$ and $\int_B$ are multidimensional (m-dimensional) integrals; therefore, an expression, for instance of the type $\int_A P_x(y)dy$ is an abbreviation of the integral

$$\int \underset{A}{\cdots} \int P_x(y_1, \ldots, y_n)dy_1 \ldots dy_n$$

$$P_{\text{лт}}=0,\ P_{\text{пр}}=1. \tag{14.9}$$

If region B equals zero, then it will be

$$P_{\text{лт}}=1,\quad P_{\text{пр}}=0. \tag{14.10}$$

Consequently, it is possible to obtain any relationship between error probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$ (any ratio $P_{\text{лт}}/P_{\text{пр}}$ value from zero to infinity) through appropriate selection of the boundaries between regions A and B.

The optimum receiver must create a boundary between regions A and B that, in a certain sense, insures a relationship between error probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$.

It follows from relationships (14.9) and (14.10) and from Figure 14.3 that, when there is a boundary change leading to a decrease in probability $P_{\text{лт}}$, there is an increase in probability $P_{\text{пр}}$ and vice versa. Therefore, it is impossible to use selection of region boundaries to provide a simultaneous minimum for /244 both $P_{\text{лт}}$ and $P_{\text{пр}}$ and the optimum receiver may minimize only a particular combination of probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$. In accordance with this, the following binary detection optimizations are the most widespread:

1. Minimum composite error probability criterion (ideal observer criterion):

$$P_{\text{ош}}=P(x_0)P_{\text{лт}}+P(x_1)P_{\text{пр}}=\min. \tag{14.11}$$

2. Minimum average risk criterion:

$$R=aP(x_0)P_{\text{лт}}+bP(x_1)P_{\text{пр}}=\min, \tag{14.12}$$

where a and b -- weight factors based on the relative danger of false alarms and signal misses (the greater the danger of false alarms compared to misses, the greater the a/b ratio selected).

Magnitude R determined from relationship (14.12) is called the average risk.

3. Minimum weighted probability criterion:

$$z = cP_{\text{лт}} + dP_{\text{пр}} = \min, \tag{14.13}$$

where c and d — weight factors based on the relative danger of false alarms and signal misses (the greater the danger of false alarms compared to misses, the greater the c/d ratio selected).

4. Neyman-Pearson criterion:

$$\left.\begin{aligned} P_{\text{пр}} &= \min \\ P_{\text{лт}} &= \text{const.} \end{aligned}\right\} \tag{14.14}$$

where

If a priori probabilities $P(x_0)$ and $P(x_1)$ of signal presence and absence are known, then it is best to use criterion 1 or 2 since these probabilities are not considered in criteria 3 and 4, i. e., all known a priori information about the signal is not used completely. Here, it is possible to use criterion 1 in those cases when false alarms and misses are equally dangerous, while criterion 2 is used in the general case.

If a priori probabilities $P(x_0)$ and $P(x_1)$ are unknown, which occurs in many cases (in radar, for example), criteria 1 and 2 may not be used. Criteria 3 and 4 have to be used in this event.

In the event criterion 3 is used, weight factors c and d must be given, because of physical formulation of the problem (here, which will become clear later, only their ratio c/d plays a role), with false-alarm probability $P_{\text{лт}}$ given when criterion 4 is used. In a majority of practical cases, it is simpler to use the corresponding selection of magnitude $P_{\text{лт}}$ rather than ratio c/d; therefore, the Neyman-Pearson criterion is used much more often than the minimum weighted error probability criterion.

It follows from comparison of relationships (14.11)--(14.13) that the /245

first two criteria from a mathematical point of view may be considered particular cases of criterion 3: criterion 1 is obtained from (14.13) if one assumes:

$$c = P(x_0),\ d = P(x_1), \tag{14.15}$$

while criterion 2, if one assumes:

$$c = aP(x_0),\ d = bP(x_1); \tag{14.16}$$

therefore, we will examine criterion (14.13) in more detail.

The expression (14.13) minimum is insured by appropriate region A and B selection, i. e., by using magnitudes $P_{\text{лт}}$ and $P_{\text{пр}}$, given constant factors c and d. Therefore, the magnitude z minimum coincides with the minimum of combination

$$z' = P_{\text{пр}} + \beta P_{\text{лт}}, \tag{14.17}$$

where

$$\beta = \frac{c}{d}. \tag{14.18}$$

From (14.6), 14.8), and (14.17), we have

$$z' = 1 - \int_A [P_{x_1}(y) - \beta P_{x_0}(y)]\, dy. \tag{14.19}$$

Consequently, the optimum receiver must provide the expression (14.19) minimum, i. e., the maximum of magnitude

$$z'' = \int_A [P_{x_1}(y) - \beta P_{x_0}(y)]\, dy. \tag{14.20}$$

The integrand is positive for some realization y(t) values and negative for others. It is evident that region A must include all values of y for which the integrand is positive and exclude every y value where the integrand is negative in order to obtain the greatest integral z" value. This denotes that this condition must be met for region A

$$P_{x_1}(y)-\beta P_{x_0}(y)>0,$$

i. e.,

$$\frac{P_{x_1}(y)}{P_{x_0}(y)}>\beta, \tag{14.21}$$

while this inequality corresponds to region B

$$\frac{P_{x_1}(y)}{P_{x_0}(y)}\leqslant\beta. \tag{14.22}$$

But, a "yes" response corresponds to falling into region A, while a "no" corresponds to falling into region B. Consequently, when condition (14.21) is met, the decision must be made in the optimum receiver that there is a signal, while the decision is that there is no signal if this condition is not met. /246 In other words, the optimum receiver must compute magnitude

$$l(y)=\frac{P_{x_1}(y)}{P_{x_0}(y)} \tag{14.23}$$

and compare it with its "threshold" β (Figure 14.4); if it turns out here that

$$l(y)>\beta, \tag{14.24}$$

the decision is that there is a signal; otherwise, the signal is assumed to be absent.

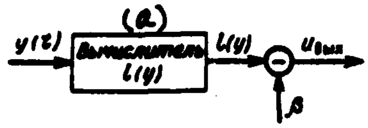


Figure 14.4. (a) — l(y) computer.

Magnitude l(y) determined from expression (14.23) is called the likelihood factor* since it equals the ratio of likelihood functions $P_{x1}(y)$ and $P_{xo}(y)$ and characterizes the likelihood of the signal present hypothesis: the greater l(y), the more likely this hypothesis.

It follows from the above examination that optimum receiver structure is the same for the first three optimizations, threshold β. values being the only difference. In the case of criterion 3 (minimum weighted probability criterion)

$$\beta=\frac{c}{d};$$

in the case of the minimum composite error probability criterion, as follows from (14.15)

$$\beta=\frac{P(x_0)}{P(x_1)}, \qquad (14.25)$$

while, when the minimum average risk criterion is used, as follows from (14.16),

$$\beta=\frac{aP(x_0)}{bP(x_1)}. \qquad (14.26)$$

*Sometimes, it also is called the likelihood ratio.

For normal white noise, from formula (4.10) we have

$$P_{x_1}(y)=\frac{1}{(\sqrt{2\pi N})^n}\,e^{-\frac{1}{N_0}\int\limits_0^T[y(t)-u_{x_1}(t)]^2\,dt}$$

and

$$P_{x_0}(y)=\frac{1}{(\sqrt{2\pi N})^n}\,e^{-\frac{1}{N_0}\int\limits_0^T[y(t)-u_{x_0}(t)]^2\,dt}.$$

Therefore, in accordance with (14.23) and (14.5a), the likelihood factor /247 equals

$$l(y)=e^{-\frac{Q}{N_0}}\,e^{\frac{2}{N_0}\int\limits_0^T y(t)u_c(t)\,dt}, \tag{14.27}$$

where

$$Q=\int\limits_0^T u_c^2(t)\,dt.$$

It follows from (14.24) and (14.27) that the optimum receiver must decide that there is a signal if this condition is met

$$\xi>U_0, \tag{14.28}$$

where

$$\left.\begin{aligned}\xi&=\frac{2}{N_0}\int\limits_0^T y(t)\,u_c(t)\,dt,\\ U_0&=\frac{Q}{N_0}+\ln\beta.\end{aligned}\right\} \tag{14.29}$$

These relationships completely coincide with relationships (5.11a)--(5.11c) obtained using different approaches [the minimum composite error probability criterion, in which β is determined from formula (14.25), was examined in Chapter 5].

The optimum receiver operating principle described above is called the likelihood factor test since the decision is made based on comparing likelihood factor l(y) with some threshold β [condition (14.24)].

It was demonstrated above that this principle supplies the optimum solution to the problem for any of the first three optimizations. It will be demonstrated below that the same principle provides an optimum solution for the Neyman-Pearson criterion as well, given proper threshold β selection. In other words, optimum receiver structure turns out to be identical for all four optimizations formulated above, the only difference being selection of threshold β magnitude.

So-called performance curves depicted (qualitatively) in Figure 14.5 are the basic optimum receiver (observer) characteristics. Each performance curve

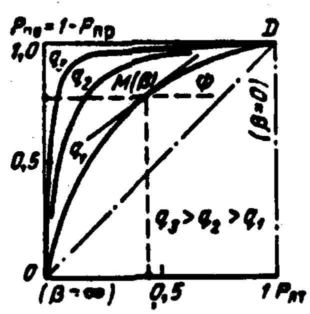


Figure 14.5

gives the dependence of correct detection probability P_{no} on false-alarm probability $P_{\text{л}}$, for given signal-to-noise power ratio q (here and in future, $q = Q/N_0$). It is not difficult to become convinced that performance curves actually must have (qualitatively) such a form, considering the following:

1. It follows from relationships (14.9) and (14.10) that, in any detector

$$P_{\text{по}}=0 \text{ where } P_{\text{лт}}=0;$$
$$P_{\text{по}}=1 \text{ where } P_{\text{лт}}=1.$$

Coordinates of points 0 and D in Figure 14.5 actually satisfy these /248
relationships.

2. For given false-alarm probability $P_{\text{лт}}$, the greater the signal-to-noise ratio q, the greater the correct detection probability must be. Consequently, the greater parameter q, the higher the location of the corresponding performance curve must be.

3. A specific threshold β value corresponds to each curve point M.

Actually, it follows from relationship (14.24) that, for a given signal-to-noise ratio q and $\beta\to\infty$, $P_{\text{лт}}$ = 0 and $P_{\text{по}}$ = 0 must be the case (since the probability of exceeding an infinitely-high threshold equals zero, both when a signal is absent and when it is present). Consequently, $\beta=\infty$. corresponds to point 0.

If $\beta=0$, then the probability of exceeding such a threshold equals unity when a signal is and is not present. Consequently, where $\beta=0$, $P_{\text{по}}=1$ and $P_{\text{лт}}=1$ must be the case, i. e., $\beta=0$. corresponds to point D (Figure 14.5). Thus, when threshold β changes from infinity to zero, point M shifts along the performance curve from the origin to point D; one fully-determinate point on the performance curve corresponds to each β value and, on the other hand, one fully-determinate threshold β value corresponds to each point M on this curve.

4. Performance curve properties noted in points 1--3 are present in both optimum and non-optimum receivers.

If the receiver is optimum (in the sense of any of the four optimizations enumerated), then this relationship also is satisfied for it

$$\operatorname{tg}\psi=\beta, \qquad (14.30)$$

where ψ — slope of the tangent to the performance curve at given point M of this curve (Figure 14.5).

Since $\beta = 0$, results at point D, then $\psi = 0$, at this point, i. e., the tangents to the performance curves are horizontal. At point O, where $\beta = \infty$,, the result is $\psi = \frac{\pi}{2}$, i. e., the tangents to the performance curves are vertical. Hence, it follows that, in optimum receivers, performance curves in their initial sector (at point O) have vertical tangents, while the tangents are horizontal at the final sector (at point D).

Relationship (14.30) validity is demonstrated in the following manner.

It follows from Figure 14.5 that

where
$$\left.\begin{aligned}\operatorname{tg}\psi=\frac{dP_{\text{п о}}}{dP_{\text{л т}}}\\ q=\text{const.}\end{aligned}\right\} \qquad (14.31)$$

For given signal-to-noise ratio value q, the performance curve under study provides dependence $P_{\text{п о}} = f(P_{\text{л т}})$for an optimum receiver. Here, its value corresponds to each of this curve's points, for instance at point M /249

$$\beta = \beta_{\text{м}}.$$

In order to find what tg ψ equals, there is a requirement, retaining the receiver as optimum, to provide an infinitely-small increment $d\beta$ to threshold β. It follows from Figure 14.5 that a threshold β change causes a change in probabilities $P_{\text{п о}}$ and $P_{\text{л т}}$, i. e.,

$$P_{\text{п о}}=f_1(\beta) \text{ and } P_{\text{л т}}=f_2(\beta), \qquad (14.32)$$

therefore

$$dP_{\text{no}} = \frac{df_1(\beta)}{d\beta}\,d\beta, \quad dP_{\text{лт}} = \frac{df_2(\beta)}{d\beta}\,d\beta,$$

and, at point $M(\beta_0)$

$$\operatorname{tg}\psi = \left(\frac{dP_{\text{no}}}{dP_{\text{лт}}}\right)_{\beta=\beta_0} = \frac{\left[\frac{df_1(\beta)}{d\beta}\right]_{\beta=\beta_0}}{\left[\frac{df_2(\beta)}{d\beta}\right]_{\beta=\beta_0}}. \tag{14.33}$$

On the other hand, it follows from (14.6), (14.8), and (14.20) that an optimum receiver provides the maximum of magnitude

$$z' = P_{\text{no}} - \beta P_{\text{лт}};$$

therefore, at point M, where $\beta = \beta_0$, the maximum of the following magnitude is provided

$$z'' = P_{\text{no}} - \beta_0 P_{\text{лт}}. \tag{14.34}$$

It follows from (14.32) and (14.34) that, when β changes, function

$$z''(\beta) = f_1(\beta) - \beta_0 f_2(\beta) \tag{14.35}$$

must have a maximum at point $\beta = \beta_0$; consequently, this condition must be met

$$\left[\frac{dz''(\beta)}{d\beta}\right]_{\beta=\beta_0} = 0,$$

i. e.,

$$\left[\frac{df_1(\beta)}{d\beta}\right]_{\beta=\beta_0} - \beta_0\left[\frac{df_2(\beta)}{d\beta}\right]_{\beta=\beta_0} = 0.$$

Considering this relationship, from (14.33) we obtain

$$\operatorname{tg}\psi=\beta_0.$$

Consequently, for a receiver optimum in the sense of any of the first three criteria [i. e., which provide the expression (14.17) minimum or the expression (14.34) maximum], the performance curves actually satisfy relationship (14.30).

Now, we will demonstrate that a receiver optimum in the sense of any of the first three criteria turns out optimum also in the sense of the Neyman-Pearson criterion, given proper threshold β selection.

This requires demonstrating that, for a given $P_{\text{л т}}$ magnitude, a receiver /250 providing the magnitude z" maximum [formula (14.34)] also provides the probability $P_{\text{п о}}$ maximum, given proper β_0 selection.

The performance curve of a receiver (optimum in the sense of the first three criteria) for some signal-to-noise ratio q_1 is depicted in Figure 14.6. Let the

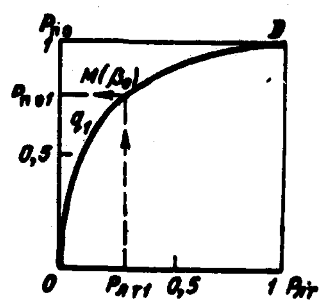


Figure 14.6

false-alarm probability be given

$$P_{\text{л т}}=P_{\text{л т}1}.$$

It follows from Figure 14.6 that fully-determinate point M on the performance curve corresponds to given q_1 and $P_{\text{л т}1}$ and, consequently, fully-determinate threshold

β and correct detection probability values:

$$\beta=\beta_0;\quad P_{\text{п о}}=P_{\text{п о1}}. \tag{14.36}$$

We will show that, for given values $q = q_1$ and $P_{\text{л т}}=P_{\text{л т1}}$, selection of a receiver having such a performance curve and such a threshold $\beta=\beta_0$, insures that the probability $P_{\text{п о}}$ maximum is obtained, i. e., satisfies the Neyman-Pearson criterion.

Each receiver type and each output threshold value (for a given receiver type) has its own region A shape (i. e., the region used as the basis for the signal/no signal decision). Selection of region A determines probability $P_{\text{п о}}$ and $P_{\text{л т}}$ magnitudes (for a given q).

Consequently,

and
$$\left.\begin{aligned}P_{\text{п о}}&=P_{\text{п о}}[A(\beta)]\\P_{\text{л т}}&=P_{\text{л т}}[A(\beta)],\end{aligned}\right\} \tag{14.37}$$

where notations $P_{\text{п о}}[A(\beta)]$ and $P_{\text{л т}}[A(\beta)]$ denote that $P_{\text{п о}}$ and $P_{\text{л т}}$ will depend on region A, which in turn will depend on threshold β.

Let region $A_1(\beta_0)$ correspond to the optimum receiver and to threshold β_0 and region $A(\beta)$ to a receiver operating on another principle (or having another threshold β value).

Since given false-alarm probability $P_{\text{л т1}}$ must be insured in all cases, then

and
$$\left.\begin{aligned}P_{\text{л т}}[A_1(\beta_0)]&=P_{\text{л т1}}\\P_{\text{л т}}[A(\beta)]&=P_{\text{л т1}}.\end{aligned}\right\} \tag{14.38}$$

Consequently, it remains to show that, for $\beta=\beta_0$ and region A of the $A_1(\beta_0)$ type, probability $P_{\text{п о}}$ will be greater than for any other region A type, i. e.,

that

$$P_{no}[A_1(\beta_0)] > P_{no}[A(\beta)]. \tag{14.39}$$

It follows from (14.34) that region $A_1(\beta_0)$ corresponds to a receiver /251 insuring the maximum of magnitude

$$z' = P_{no} - \beta_0 P_{\text{лт}},$$

i. e.,

$$P_{no}[A_1(\beta_0)] - \beta_0 P_{\text{лт}}[A_1(\beta_0)] > P_{no}[A(\beta)] - \beta_0 P_{\text{лт}}[A(\beta)]. \tag{14.40}$$

But, it follows from (14.38) that

$$P_{\text{лт}}[A_1(\beta_0)] = P_{\text{лт}}[A(\beta)];$$

therefore, (14.40) coincides with (14.39), which must be demonstrated.

Consequently, a receiver optimum in the sense of the first three criteria also is optimum in the sense of the Neyman-Pearson criterion if threshold β is selected with respect to $P_{\text{лт}}$ and q_1, as shown in Figure 14.6.

Thus, for any of the aforementioned optimizations, the receiver has the identical structure insuring computation of likelihood factor $l(y)$ and its comparison with threshold β (Figure 14.4). Only threshold β magnitude will depend on the optimization accepted.

Threshold β for criteria 1, 2, and 3 is determined from formulas (14.25), (14.26), and (14.18), respectively, and will not depend on the signal-to-noise ratio. In the case of the Neyman-Pearson criterion, β is determined from performance curves (Figure 14.5) for given $P_{\text{лт}}$ and q values and, consequently, will depend on the signal-to-noise ratio.

14.4 Detection Characteristics. Threshold Signal

Detection characteristics provide the dependence of incorrect (or correct) detection probability on the signal-to-noise ratio and make it possible to determine the threshold signal (signal-to-noise ratio threshold value), i. e., the minimum signal insuring given detection quality. Detection characteristics may be plotted from performance curves and their form will depend on the optimization selected.

In the case of the minimum composite probability criterion (ideal observer criterion), composite error probability $P_{ош}$ determined from expression (14.11) characterizes detection quality. Therefore, the following dependence is called the detection characteristic

$$P_{ош} = f_1(q).$$

It may be plotted from performance curves (Figure 14.5) in the following manner:

1. We determine threshold β from formula (14.25).

2. Using the graphical approach, we will find the points (one each per curve) on the Figure 14.5 performance curves at which the slope of the curve equals β , i. e. for which

$$\operatorname{tg}\psi = \beta.$$

3. For every point found, i. e., for each value of q, we determine /252
magnitudes $P_{лт}$ and $P_{пр}$ and we compute the corresponding composite probability $P_{ош}$ values from formula (14.11).

4. We plot the dependence of $P_{ош}$ on q from the values found.

Thus, for example, when $\beta = 1$, the detection characteristic of a fluctuating signal has the form depicted in Figure 9.8.

In the case of criteria 2 and 3, the following dependencies, respectively, are called detection characteristics

$$R=f_2(q) \quad \text{and} \quad z=f_3(q).$$

The methodology for plotting them is analogous to that described above, with the exception that threshold β is determined accordingly from formula (14.26) or (14.18), while magnitudes R and z are computed from formulas (14.12) and (14.13).

In the case of the Neyman-Pearson criterion, this dependence is called the detection characteristic

$$P_{\text{по}}=f_4(q) \text{ where } P_{\text{лт}}=\text{const.}$$

It will be plotted from performance characteristics (Figure 14.5) in the following manner:

1. A vertical line is drawn through point ($P_{\text{лт}}$, 0), where $P_{\text{лт}}$ — given false-alarm probability, and the points where it intersects the performance curves will be found.

A $P_{\text{по}}$ magnitude will be found for each of these points, namely for every q value.

2. Dependence $P_{\text{по}}$ = $f_4(q)$ will be plotted from the values found.

Such a characteristic computed from the Figure 9.6 curves for $P_{\text{лт}}$ = 0.2 is depicted in Figure 14.7.

Signal-to-noise ratio $q_{\text{мин}}$, insuring the given correct detection probability is determined from the detection characteristic found.

Fully-determinate point M in a family of performance curves (Figure 14.5) corresponds to the $q_{\text{мин}}$ value and given probability $P_{\text{лт}}$ found. The performance curve slope at this point determines the requisite threshold β magnitude:

$$\beta=\text{tg}\,\psi.$$

14.5 Comparison of the Statistical Hypothesis Testing Approach to the Inverse Probability Approach

In accordance with the inverse probability approach presented in Parts II and III, the signal/no signal decision is based on comparison of inverse probabilities $P_y(x_1)$ and $P_y(x_0)$ of signal presence and absence, respectively. If the /253 maximum inverse probability criterion is used, then the decision that there is a signal is made if

$$P_y(x_1) > P_y(x_0), \tag{14.41}$$

and that there is no signal in the opposite case. Here, minimum composite error probability P_{om} is insured.

It also was pointed out that, in those cases when the danger of false alarms and signal misses is not identical, the decision that there is a signal must be made when this condition is met

$$P_y(x_1) > \eta P_y(x_0), \tag{14.42}$$

where η -- some weight factor selected that is greater, the greater the danger of false alarms compared to signal misses.

It is not difficult to become convinced that these criteria completely coincide with the minimum composite error probability and minimum average risk criteria presented in this chapter.

Actually, considering that

$$\left.\begin{aligned} P_y(x_1) &= kP(x_1)\,P_{x_1}(y) \\ P_y(x_0) &= kP(x_0)\,P_{x_0}(y), \end{aligned}\right\} \tag{14.43}$$

and

condition (14.42) may be written in the form

$$
\left.\begin{array}{l} l(y) > \beta, \\ \text{where} \quad l(y) = \dfrac{P_{x_1}(y)}{P_{x_0}(y)} \text{ и } \beta = \eta \dfrac{P(x_0)}{P(x_1)}. \end{array}\right\} \tag{14.44}
$$

These relationships completely coincide with expressions (14.23), (14.24), and (14.26) if you assume

$$\eta = \frac{a}{b}. \tag{14.45}$$

Consequently, criterion (14.42), which may be called a weighted maximum inverse probability criterion, completely coincides with the minimum average risk criterion, i. e., insures minimum average risk.

If you assume

$$\eta = 1, \tag{14.46}$$

then condition (14.42) corresponds to the maximum inverse probability criterion and, in formulas (14.44), the following should be assumed

$$\beta = \frac{P(x_0)}{P(x_1)}.$$

These results completely coincide with the expressions obtained from the minimum composite error probability criterion (ideal observer criterion). /254 Consequently, the maximum inverse probability criterion coincides with the minimum composite error probability criterion (ideal observer criterion).

Statistical criteria 3 and 4 are used, as indicated in § 14.3, in those cases when a priori probabilities $P(x_0)$ and $P(x_1)$ are unknown and, consequently, inverse probabilities $P_y(x_1)$ and $P_y(x_0)$ also are unknown [see formulas (14.43)]. Therefore, criteria 3 and 4 cannot be equated with criteria based on comparison of inverse probabilities. However, as was demonstrated in § 14.3, the structure of the optimum

receiver for criteria 3 and 4 is identical to that used with criteria 1 and 2--the only difference is threshold β magnitude. Therefore, during optimum detector design and investigation of its properties, it is relatively immaterial which criterion is used.

The following two criteria are the most widespread due to their simplicity and clarity:

1. Minimum composite error probability criterion (ideal observer criterion, maximum inverse probability criterion).

2. Neyman-Pearson criterion.

The first criterion is used more often in communications systems, while the second more often in radar (since a priori probabilities $P(x_1)$ and $P(x_0)$ in many cases are knowns in communications systems, while as a rule they are unknowns in radar).

14.6 Computation of the Likelihood Factor for a Random-Parameter Signal

It was demonstrated in § 14.3 that, in the case of a precisely-known signal, the optimum receiver (detector) must compute likelihood factor $l(y)$ and compare it with threshold β . Here, the likelihood factor is determined from expression (14.23). We now will explain the special features arising if a signal has parasitic random parameters, i. e.,

$$y(t)=u_x(\alpha_1,\ldots,\alpha_n;\ t)+u_m(t).$$

Here, in accordance with (8.11) and (8.12), we have (assuming that message x has no statistical link with the parasitic parameters):

$$\left.\begin{aligned} P_y(x_1) &= k\int\limits_{A_{\alpha_1}}\ldots\int\limits_{A_{\alpha_n}} P(x_1)\,P(\alpha_1,\ldots,\alpha_n)\times \\ &\quad\times P_{x_1,\alpha_1,\ldots,\alpha_n}(y)\,d\alpha_1\ldots d\alpha_n; \\ P_y(x_0) &= kP(x_0)\,P_{x_0}(y). \end{aligned}\right\} \qquad (14.47)$$

Therefore, condition (14.41), corresponding to the maximum composite error probability criterion (or the maximum inverse probability criterion), may be written in the following form:

$$l(y) > \beta, \tag{14.48}$$

where /255

$$l(y) = \frac{\int\limits_{A_{\alpha_1}} \cdots \int\limits_{A_{\alpha_n}} P(\alpha_1, \ldots, \alpha_n) P_{x_1, \alpha_1, \ldots, \alpha_n}(y)\, d\alpha_1 \ldots d\alpha_n}{P_{x_0}(y)}; \tag{14.49}$$

$$\beta = \frac{P(x_0)}{P(x_1)}. \tag{14.50}$$

Comparing expressions (14.48)--(14.50) with the corresponding expressions for a precisely-known signal, i. e., with formulas (14.23)--(14.25), it is not difficult to become convinced that they coincide, with the exception that, when computing likelihood factor l(y), the value of likelihood function $P_{x_1, \alpha_1, \ldots, \alpha_n}(y)$, averaged considering a priori distribution $P(\alpha_1, \ldots, \alpha_n)$ of all parasitic signal parameters, will be placed in the numerator.

It is possible also to write expression (14.49) in the form

$$l(y) = \int\limits_{A_{\alpha_1}} \cdots \int\limits_{A_{\alpha_n}} P(\alpha_1, \ldots, \alpha_n)\, l_{\alpha_{1n}}(y)\, d\alpha_1 \ldots d\alpha_n, \tag{14.51}$$

where

$$l_{\alpha_{1n}}(y) = \frac{P_{x_1, \alpha_1, \ldots, \alpha_n}(y)}{P_{x_0}(y)} \tag{14.52}$$

is the likelihood factor computed in the assumption that all parasitic parameters $\alpha_1, \ldots, \alpha_n$ of the signal are known, i. e., that the signal is precisely known.

It follows from relationships (14.51) and (14.52) that likelihood factor l(y) corresponding to a signal with parasitic parameters may be obtained through statistical averaging of likelihood factor $l_{\alpha 1n}(y)$ (found in the assumption that the signal is precisely known) relative to all parasitic parameters $\alpha_1, \ldots, \alpha_n$.

Formulas (14.48)--(14.52) were derived above relative to the minimum composite error probability criterion. All these formulas remain valid for the remaining criteria, as in the case of a precisely-known signal, with the exception that threshold β must be determined from formulas corresponding to these criteria [formulas (14.18), (14.26), and so on], rather than from formula (14.50).

Optimum detector structure and error probabilities $P_{\text{лт}}$, $P_{\text{пр}}$ and $P_{\text{ош}}$ for various signal types may be found on the basis of relationships (14.18)--(14.50).

However, as was demonstrated in § 14.5, an approach based upon computation of likelihood factor l(y) and its comparison with threshold β, fully equates to the approach based upon computation and comparison of inverse probabilities $P_y(x_1)$ and $P_y(x_0)$ presented in Part III of the book. Therefore, use of the likelihood factor approach to analysis of specific cases, already examined in Part III (detection of a random-phase, fluctuating, and other signal types), would only be a repeat of these results and is senseless.

Mention only should be made that some mathematical computations may be /256
simplified (compared to those presented in Part III of the book) if the property of the optimum detector performance curves characterized by relationship (14.30) is used.

It follows from relationships (14.30) and (14.31) that, for the optimum detector

$$\frac{dP_{\text{по}}(\beta)}{dP_{\text{лт}}(\beta)} = \beta, \qquad (14.53)$$

where designations $P_{\text{по}}(\beta)$ and $P_{\text{лт}}(\beta)$ indicate that correct detection probability $P_{\text{по}}$ and false-alarm probability $P_{\text{лт}}$ here are considered a threshold β function (Figure 14.5).

It follows from formula (14.53) and Figure 14.5 that

$$P_{\text{по}}(\beta) = \int_{\infty}^{\beta} \beta \left[\frac{dP_{\text{лт}}(\beta)}{d\beta}\right] d\beta \qquad (14.54)$$

[in the case examined, integration occurs with respect to curve OM (Figure 14.5)

and variable β changes from infinity (at point 0) to value β, corresponding to point M at which magnitude $P_{n\,o}(\beta)$ is determined].

Formula (14.54) makes it possible to simplify computation of correct detection probability $P_{n\,o}$ significantly since false-alarm probability $P_{n\,T}$ included in it is determined when only noise (without signal) is active, while determination of magnitude $P_{n\,o}$ without using expression (14.54) requires examination of a case where signal-plus-noise is active.

We will examine detection of a random-phase signal to illustrate these postulations. The following results were obtained for this case in Chapter 9:

$$P_{n\,T} = e^{-(N_0/4Q_0)z^2}, \qquad (9.34)$$

where

$$\ln I_0(z) = \frac{Q_0}{N_0} + \ln\frac{P(0)}{P(a_0)}; \qquad (9.35)$$

here P(0) and P(a_0) -- a priori probabilities of signal absence and presence, i. e.,

$$P(0) = P(x_0) \quad \text{and} \quad P(a_0) = P(x_1).$$

Since formula (9.35) was obtained from the maximum inverse probability criterion (or the minimum composite error probability, which is the same thing), then relationship (14.25) is valid, i. e.,

$$\frac{P(0)}{P(a_0)} = \frac{P(x_0)}{P(x_1)} = \beta.$$

Therefore, it is possible to write formula (9.35) in the following form: /257

where
$$\left.\begin{aligned} I_0(z) &= e^{Q_0/N_0}\beta, \\ \beta &= e^{-Q_0/N_0} I_0(z). \end{aligned}\right\} \qquad (14.55)$$

It follows from these relationships that a one-to-one link exists between z and β; therefore, formula (9.34), in essence, expresses dependence $P_{\text{лт}}(\beta)$ of interest to us and expression (14.54) may be written in the following form:

$$P_{\text{по}} = \int\limits_{\infty}^{z} e^{-Q_0/N_0} I_0(z) \frac{dP_{\text{лт}}(z)}{dz} dz. \tag{14.56}$$

But, if follows from (9.34) that

$$\frac{dP_{\text{лт}}}{dz} = -\frac{N_0}{2Q_0} z e^{-(N_0/4Q_0)z^2};$$

therefore, from (14.56) we have

$$P_{\text{по}} = 1 - P_{\text{пр}} = \frac{N_0}{2Q_0} e^{-Q_0/N_0} \int\limits_{z}^{\infty} x e^{-(N_0/4Q_0)x^2} I_0(x) dx. \tag{14.57}$$

This expression fully coincides with formula (9.36) obtained in Chapter 9 by another, more-unwieldy approach.

CHAPTER FIFTEEN /257

SEQUENTIAL DETECTION

15.1 General Comments

The theory of optimum reception methods presented above is based on the fact that signal (or message) detection or reproduction must occur during predetermined observation time T. However, in many cases, better results may be obtained if the observation duration is not predetermined (i. e., prior to beginning the observation) and the problem of when observation should cease is solved during the observation process itself, depending on results obtained.

Actually, in some sequences, noise oscillation realizations may turn out to be so favorable that reliable detection of a signal and reproduction of its parameters may occur much faster than in other sequences, when noise realizations are less favorable. Therefore, if observation duration T is not predetermined, it is possible to obtain a significant savings in observation time on the average (for many sequences).

A case where observation duration T is not predetermined, but is determined /258 by the progress of the experiment itself, is called sequential observation (sequential analysis). As opposed to this, observation with a predetermined duration is called simple or classic observation.

It follows from the aforementioned that duration T of the time of observation during sequential observation is a random magnitude changing from one sequence to another.

In those sequences in which noise realizations turn out to be favorable, resultant observation time is less than its mean value $\bar{T}$.

On the other hand, time T may turn out to be significantly greater than $\bar{T}$ when noise realizations are unfavorable.

The main advantage of sequential observation over simple (classic) observation is the decrease in observation time mean value $\bar{T}$ (it is evident that $\bar{T}$ = T during simple observation). Its main drawback is the random nature of the observation time and the attendant possibility of situations where T turns out to be much greater than $\bar{T}$.

Wald for the first time did a mathematical study of the sequential observation (sequential analysis) process and published on this subject in 1947 [29]. In subsequent years, Middleton, Busgang, Blasbalg, Bashirinov, Fleyshman, and others successfully used the Wald vehicle to solve the problem of optimum sequential detection of a signal on a noise background [16, 78, 85, 86, and 121]. A brief rundown on the basic results obtained in these works is presented in this chapter.

15.2 Sequential Detector Operating Principle

We will examine sequential binary detection of a signal. As a result of observation, the decision must be "yes" (signal) or "no" (no signal, only noise).

The observation process is divided into a series of sequential intervals (steps) of sufficiently-short duration Δt (Figure 15.1). The "observer" (receiver) in the first observation step has signal-plus-noise realization $y_{o1}(t)$ at its disposal (Figure 15.2a). The likelihood factor is computed from formula (14.23) based on this realization

$$l_1(y_{o1}) = \frac{P_{x_1}(y_{o1})}{P_{x_0}(y_{o1})}. \qquad (15.1)$$

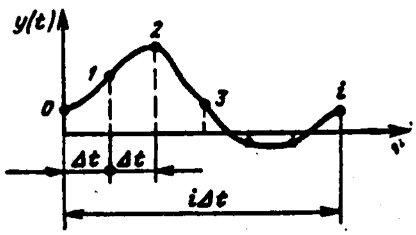


Figure 15.1

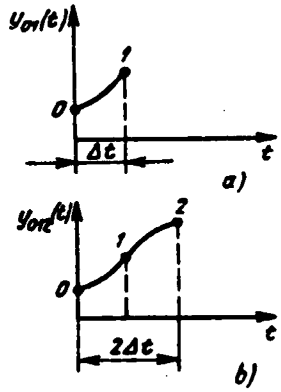


Figure 15.2

If a "yes" or "no" decision must be made by the end of interval Δt, then the likelihood factor should be compared with some threshold β (as is the case in a simple analysis) and, if this threshold is overrun, the response is "yes." Otherwise, the response is "no." However, during <u>sequential</u> detection, the /259 time the experiment ends (i. e., the moment the decision is made) is not predetermined; therefore, three, rather than two, different answers are possible at the end of interval Δt:

1. "Yes" (signal).

2. "No" (no signal)

3. It is impossible as yet to decide ("yes" or "no") with sufficient reliability and, consequently, the observation must be extended, i. e., move to the next observation step.

Therefore, two thresholds β_1 and β_2 $(\beta_2 > \beta_1)$, rather than one, are established during sequential detection.

If it turns out that

$$l_1(y_{01}) > \beta_2, \tag{15.2a}$$

the decision is "yes" (signal) and, consequently, observation ceases; if

$$l_1(y_{01}) < \beta_1, \tag{15.2b}$$

then the decision is "no" (no signal) and, consequently, observation also ceases.

Finally, if it turns out that

$$\beta_1 \leqslant l_1(y_{01}) \leqslant \beta_2, \tag{15.2c}$$

then no decision is made and observation continues, i. e., a transition is made to the second step.

In the second step, the "observer" (receiver) already has at its disposal realization $y_{012}(t)$, which corresponds to interval $2\Delta t$ (Figure 15.2b) and the likelihood factor is computed for it:

$$l_2(y_{012}) = \frac{P_{s1}(y_{012})}{P_{s0}(y_{012})}. \tag{15.3a}$$

Magnitude $l(y_{012})$ again is compared with thresholds β_1 and β_2 and the decision is "yes" if

$$l_2(y_{012}) > \beta_2,$$

or is "no" if /260

$$l_2(y_{012}) < \beta_1.$$

If it turns out that

$$\beta_1 \leqslant l_2(y_{012}) \leqslant \beta_2, \tag{15.3b}$$

then no decision is made and there is a transition to the next, third, observation step.

Next, the process continues in an analogous manner until, at some step n

likelihood factor $l_n(y_{01...n})$ finally turns out not to be greater than β_2 or less than β_1 and, consequently, the appropriate decision will be made ("yes" or "no").

It may seem at first glance that, in several situations, sequential observation may last an infinitely-long time. However, as Wald demonstrated, the probability that the test will end (during a finite time) equals unity for a very broad class of realization y_i distributions.

Step n in which observation (analysis) ceases is called a finite step. Evidently, n is a random magnitude changing from one sequence (concluded observation) to another. Accordingly, also random is composite observation time

$$T = n\Delta t. \tag{15.4}$$

It follows from what has been stated that, during sequential observation, the space of all possible realizations y(t) is divided into three adjacent, but

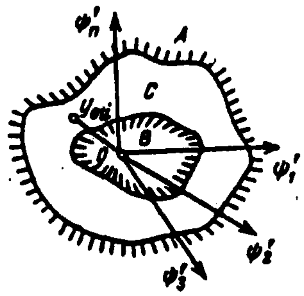


Figure 15.3

non-overlapping, regions A, B, and C (Figure 15.3). If, at the i-th step, realization $y_{01...i}(t)$ falls into regions A or B, then the appropriate "yes" or "no" decision is made and, consequently, observation ceases. If this realization $y_{01...i}(t)$ corresponds to intermediate region C, then no decision is made and observation continues.

As was the case in simple observation, two types of erroneous decisions are possible—false alarms ("yes" when there is no signal) and signal misses ("no"

when there is a signal), with probabilities $P_{лт}$ and $P_{пр}$, respectively. Regions A, B, and C in the optimum receiver must be divided so that the best, in a certain sense, solution to the detection problem is insured.

It is possible to characterize sequential detection quality by means of a linear combination of detection error probability and mean observation time, having the form /261

$$\left.\begin{aligned} R &= c_1 P(0) P_{лт} + c_2 P(C) P_{пр} + c_3 \bar{T}, \\ \text{where} \quad \bar{T} &= P(0)\overline{T(0)} + P(C)\overline{T(a_1)}. \end{aligned}\right\} \quad (15.5)$$

Here, as was the case previously, P(0) and P(C) -- a priori probabilities of signal absence and presence, respectively; $\bar{T}$ -- mean (unconditional) observation time value; T(0) -- mean observation time given that there is no signal (i. e., mean duration of those sequences in which there is no signal); $T(a_1)$ -- mean observation time given a signal with amplitude a_1 (i. e., mean duration of those sequences in which there is a signal with amplitude a_1); c_1, c_2, and c_3 -- some weight factors considering the relative danger of false alarms, signal misses, and greater mean observation duration.

Evidently, the lower the R value, the better the receiver.

A type (15.5) criterion differs from corresponding simple (not sequential) detection optimizations [see (14.12), for example] only in that, here, receiver quality, along with error probabilities, is characterized also by mean observation duration $\bar{T}$.

Wald and Wolfowitz demonstrated that, regardless of given error probabilities $P_{лт}$ and $P_{пр}$, weights c_1, c_2, and c_3, and a priori probabilities P(0) and P(C), no observation method provides mean observation duration values T(0) and $T(a_1)$ lower than does observation based on the aforementioned sequential computation of likelihood factors $l_1(y_{01})$, $l_2(y_{012})$ and so forth and their comparison with thresholds β_1 and β_2.

Here, thresholds β_1 and β_2 must be determined from formulas

$$\beta_1 = \frac{P_{\text{пр}}(a_1)}{1 - P_{\text{лт}}}; \quad \beta_2 = \frac{1 - P_{\text{пр}}(a_1)}{P_{\text{лт}}} \qquad (15.6)$$

(these formulas are valid when $P_{\text{пр}}(a_1) \leqslant 0{,}5$ and $P_{\text{лт}} \leqslant 0{,}5$, which usually is the case).

Sequential detection determined in this manner is called optimum sequential detection.

It follows from what has been said such detection insures receipt of minimum mean durations $\overline{T(0)}$ and $\overline{T(a_1)}$ and, consequently, minimum mean observation time $\overline{T}$ as well for any error probability $P_{\text{лт}}$ and $P_{\text{пр}}$ values, any a priori probability P(0) and P(C) values, and any weight factors c_1, c_2, and c_3.

It should be noted that Wald obtained formulas (15.6) with a simplified assumption that, at the end of a test, likelihood factor $l_n(y_{01. \ . \ .n})$ turns out precisely to equal threshold β_2 or β_1, i. e., so that there is no so-called "excursion outside the boundaries."

If $\Delta t \rightarrow 0$, then this assumption is completely valid. Actually, if the test ends in the n-th step, then this signifies that, in the (n—1)-th step, /262 it still has not concluded and, consequently, in the (n—1)-th step, magnitude $l_{n-1}(y_{01. \ . \ .n-1})$ still has not exceeded the boundaries set by thresholds β_1 and β_2.

On the other hand, the test is over in the n-th step and, consequently, in this step magnitude $l_n(y_{01. \ . \ .n})$ must exceed one of the boundaries determined by thresholds β_1 and β_2.

But, where $\Delta t \rightarrow 0$, the likelihood factor in the n-th step is identical to that in the (n—1)-th step. Therefore, if it does not overrun the boundaries in the (n—1)-step, then it may not overrun either boundary in the n-th step either--in this final step, it only coincides with one of these boundaries.

If interval Δt is finite, then the result in the n-th step may be $l_n > \beta_2$, or $l_n < \beta_1$, i. e., a transition outside the boundaries may occur. However, as demonstrated in several works, where $n \gg 1$ (which usually is the case), formulas (15.6) and the approximate relationships Wald obtained on their basis essentially are accurate enough.

Observation step duration Δt selection is based on the following circumstances.

Let Δt_k be that minimum interval between adjacent realization y(t) sample values in which these values may be considered statistically independent. Then, if $\Delta t > \Delta t_k$ is the selection, mean observation time $\bar{T}$ increases when interval Δt increases. Actually, in this case, a Δt increase considerably increases information about observed process y(t) and, in some cases, a sufficiently-reliable decision may be obtained also with less information, i. e., even prior to conclusion of a given interval Δt. Therefore, in several cases, an interval Δt decrease makes it possible to speed up the decision-making process and, consequently, must lead to a decrease in mean observation duration $\bar{T}$. Consequently, it is not a good idea to select an interval Δt magnitude exceeding Δt_k.

If $\Delta t \ll \Delta t_k$ is selected, then a further interval Δt decrease does not provide a significant mean observation duration decrease. Actually, in this case, all process y(t) sample values within interval Δt have such a strong mutual statistical link that each ordinate $y(t_1)$ taken within this interval essentially completely characterizes entire realization y(t) in this interval. Therefore, further breakdown of this interval Δt into a series of smaller intervals may not provide any significant improvement in detection quality and (in the case of a discrete sample) will lead only to excess system complication (due to the increase in the number of observation steps).

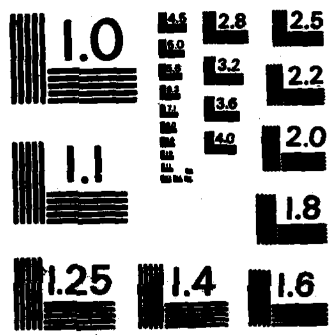
In light of the aforementioned circumstances, the usual selection is /263

$$\Delta t \approx \Delta t_k. \tag{15.7}$$

Here, for the purposes of computation and design simplification, systems usually are limited to consideration within each interval Δt only of one ordinate

AD-A120 899 THEORY OF OPTIMUM RADIO RECEPTION METHODS IN RANDOM NOISE(U) FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OH L S GUTKIN 24 SEP 82 FTD-ID(RS)T-0784-82 5/7

UNCLASSIFIED F/G 9/4 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

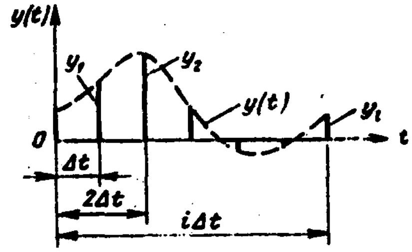


Figure 15.4

y_i, i. e., replacement of the analog sample (Figure 15.1) by a discrete sample (Figure 15.4).

If interval Δt here is selected in accordance with relationship (15.7), then this replacement of an analog sample by a discrete sample, on the one hand, does not provide a significant deterioration in detection quality while, on the other hand, it makes it possible to consider all sample values $y_1, y_2, \ldots, y_n$ statistically independent, which will lead to very-significant simplification of the analysis.

15.3 Fundamental Relationships During Optimum Sequential Detection

From the point of view of mathematical statistics, the relationships presented below correspond to the following problem formulation.

Random sample values $y_1, y_2, \ldots, y_i$ are observed sequentially. The multidimensional law of distribution of these values is fully known (for any number i of these values), except for single unknown parameter θ and is designated

$$P_\theta(y) = W(y_1, y_2, \ldots, y_i; \theta). \qquad (15.8)$$

Based on observation, the hypothesis that $\theta < \theta_0$, must be tested relative to alternative (incompatible) hypothesis H_1 that $\theta \geqslant \theta_1$ (where $\theta_1 > \theta_0$).

In order for this problem formulation to correspond to the signal detection problem, it evidently is necessary to meet the following conditions:

1. The multidimensional law of distribution of signal-plus-noise y(t) must have only one unknown parameter θ .

2. Parameter θ must be such that case $\theta \leqslant \theta_0$ unequivocally means no signal, while condition $\theta > \theta_1$, also unequivocally must correspond to signal presence.

It is not difficult to become convinced that the second condition will be satisfied if θ represents signal "strength" (amplitude, amplitude mean statistical value, energy, energy mean value, signal-to-noise power ratio, and so on) determined so that it equals zero when there is no signal $(\theta = \theta_0 = 0)$, and it is no less than some predetermined magnitude ($\theta \geqslant \theta_1$) when there is a signal.

We will examine several examples.

1. Signal a cos $(\omega t + \varphi)$ is precisely known except for amplitude a, which may have only two values: a = 0 (no signal) and a = a_1 (signal).

In this case, signal-plus-noise /264

$$y(t) = u_c(t) + u_m(t)$$

has only one unknown parameter--signal amplitude* and, if one selects

$$\theta = a; \quad \theta_1 = a_1, \tag{15.9}$$

then the second condition also will be met.

2. Signal a cos $(\omega t + \varphi)$ is known, with the exception of amplitude a

*Here and henceforth, the assumption everywhere is that the noise law of distribution is completely known.

and phase ϕ, while phase ϕ is random and equally probable in the 0 to 2π range.

In this case, it is possible to sequentially analyze the oscillation at the output of the amplitude detector connected beyond the optimum linear filter, rather than at receiver input. Here, detector output voltage values are used as sample values $y_1, y_2, \ldots, y_i$ and amplitude a is the only unknown parameter in the distribution of these values.

If the signal, when present, has the only possible known amplitude a_1 value, then it is possible to assume

$$\theta = a; \quad \theta_1 = a_1. \tag{15.10}$$

3. If, in the preceding two examples, incoming signal amplitude a is unknown, some value a_1 of this amplitude, for which miss probability $P_{sp}(a_1)$ must be given, is given prior to the observation, with the assumption being

$$\theta = a; \quad \theta_1 = a_1. \tag{15.11}$$

Composite sequential detection stemming from this assumption will be optimum, i. e., will provide minimum mean observation durations T(0) and T(a), if the amplitude of the signal arriving at receiver input actually does equal selected magnitude a_1. If not, receiver action then will not be optimum. Therefore, usually the minimum signal amplitude value at which the given detection ... must be provided is selected as a_1. This then guarantees that receiver action ... e optimum for the minimum (reliably-observable) signal.

Receiver action still will not be optimum when signal amplitude is increased beyond value a_1, but this is not so dangerous since a signal strength increase will lead to a signal miss probability decrease, even if some optimality is lacking in receiver action.

4. If what is present is a noise-type signal with a completely-known law of distribution, then the only unknown parameter of the distribution of /265

sample values (y_1, . . ., y_i) is extant signal voltage value $U_{эф}$ and it is possible to assume:

$$\theta = U_{эф}; \quad \theta_1 = U_{1эф}, \tag{15.12}$$

where

$$U_{эф} = \sqrt{\overline{u_c^2}}.$$

If magnitude $U_{эф}$ is not known beforehand, then, stemming from the same circumstances as in the preceding example, the minimum extant signal value at which given error probabilities $P_{пр}$ and $P_{лт}$ must be supplied are selected as $U_{1эф}$.

It is clear from what has been presented that, in sequential detection, θ is the parameter characterizing signal strength. Here, $\theta = 0$ corresponds to no signal, while $\theta = \theta_1$ corresponds to presence of a signal with the anticipated strength.

In the general case, strength θ of an actually-arriving signal is unknown; therefore, θ_1 is called anticipated strength value, i. e., that value determining optimal detection. As opposed to this, θ denotes that signal strength value which actually occurs during a given detection sequence and it may equal either zero, or θ_1, or (in the general case) any other value.

Prior to the start of the observation, detection error probabilities $P_{лт}$ and $P_{пр}(\theta_1)$ are given. Here, the symbol $P_{пр}(\theta_1)$ signifies that the given miss probability must be insured in those cases when signal strength equals the anticipated magnitude, i. e., θ_1.

Thresholds β_1 and β_2 determined in accordance with formulas (15.6) are selected from given $P_{лт}$ and $P_{пр}(\theta_1)$ magnitudes, namely

$$\beta_1 = \frac{P_{пр}(\theta_1)}{1 - P_{лт}}; \quad \beta_2 = \frac{1 - P_{пр}(\theta_1)}{P_{лт}}. \tag{15.13}$$

Here, $P_{лт} \leqslant 0,5$ and $P_{пр}(\theta_1) \leqslant 0,5$.

The following are basic detector (receiver) characteristics determining the quality of its work and subject to computation:

1. Dependence

$$\overline{T(\theta)}=\Delta t\cdot\overline{n(\theta)}, \tag{15.14}$$

where $\overline{T(\theta)}$ and $\overline{n(\theta)}$ -- mean (statistical) values of observation time T and sample size n, determined for the condition that input signal strength equals θ.

Since, in accordance with (15.14), a very-simple one-to-one link exists between $\overline{T(\theta)}$ and $\overline{n(\theta)}$, for brevity in future we will examine only mean sample size $\overline{n(\theta)}$.

Particular but important function $\overline{n(\theta)}$ values are $\overline{n(0)}$ and $\overline{n(\theta_1)}$, /266 i. e., mean sample values for signal absence and presence of a signal of anticipated strength θ_1, respectively.

A typical $\overline{n(\theta)}$ characteristic is depicted in Figure 15.5. It has a maximum

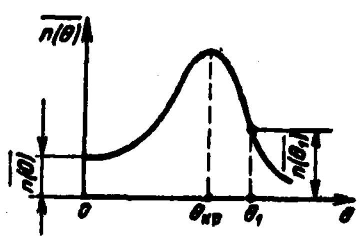


Figure 15.5

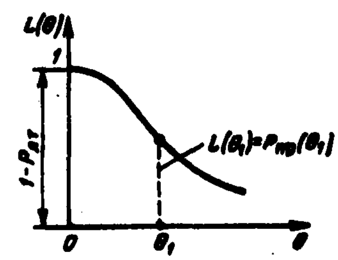


Figure 15.6

at some critical signal strength value $\theta_{кр}$, while

$$0<\theta_{кр}<\theta_1.$$

The following explains the presence of this maximum. Given sufficiently-high signal strength ($\theta > \theta_1$), the test ends relatively quickly on the average since, here, the probability is high that the total output effect of upper threshold β_2 will be exceeded.

When there is no signal (θ = 0), the test also ceases relatively quickly on the average since, here, the probability is relatively high that the effect of noise alone will be less than lower threshold β_1.

If there is a signal ($\theta \neq 0$) but its strength is relatively low ($\theta < \theta_1$), then the oscillation at receiver output [l(y) computer] will be found a majority of the time between thresholds β_1 and β_2 and a relatively-long time will pass until this oscillation turns out to be below threshold β_1 or above threshold β_2.

2. <u>Dependence</u> L(θ), called the function operational characteristic. Here, L(θ) is the probability of acceptance when the test of no-signal hypothesis H_o ceases, given that input signal strength equals θ . Since, following test cessation, it is mandatory to accept hypothesis H_o or H_1, then 1—L(θ) is the probability of accepting (following test cessation) signal hypothesis H_1, also determined given that signal strength equals θ .

It follows from these determinations that

$$\left.\begin{aligned} L(\theta_1) &= P_{\text{пр}}(\theta_1); \\ 1 - L(0) &= P_{\text{лт}}. \end{aligned}\right\} \tag{15.15}$$

Typical characteristic L(θ) is depicted in Figure 15.6. Since, for /267
$\theta \neq 0$, magnitude L(θ) is nothing but the probability of missing a signal with strength θ , then it is natural that L(θ) monotonically decreases when θ increases.

It follows from the aforementioned that the detector provides optimum results only when θ = 0 and $\theta = \theta_1$ [provides minimum $\overline{n(0)}$ and $\overline{n(\theta_1)}$ values for $P_{\text{лт}}$ and $P_{\text{пр}}(\theta_1)$]. However, it is evident from the Figure 15.5 and 15.6 curves

that, when $\theta > \theta_1$, detection quality monotonically improves with a rise in θ (mean observation time and signal miss probability decrease). Therefore, if, as demonstrated above, θ_1 corresponds to the minimum signal strength value, an excursion from the optimum detection mode occurring when $\theta > \theta_1$ is not dangerous.

Since sample size $n(\theta)$ is a random value changing from one sequence to another, it is important to know not only mean sample size value $\overline{n(\theta)}$ but the magnitude $n(\theta)$ law of distribution as well. However, to date, there has been no success in finding sufficiently-general relationships for computation of this law of distribution and most of the time, if not always, there is success in computing only variances $\sigma^2_{n(0)}$ and $\sigma^2_{n(\theta_1)}$ of magnitudes $n(0)$ and $n(\theta_1)$.

A brief description of the methodology for computing characteristics $n(\theta)$, $L(\theta)$, and dispersions $\sigma^2_{n(0)}$ and $\sigma^2_{n(\theta_1)}$, when sample values $y_1, y_2, \ldots, y_i$ are statistically independent, follows.

Function $L(\theta)$ is determined from the formula

$$L(\theta)=\frac{\beta_2^h-1}{\beta_2^h-\beta_1^h}, \tag{15.16}$$

where h -- supplemental parameter determined from functional equation

$$\int_{-\infty}^{\infty}\left[\frac{W(y_i,\theta_1)}{W(y_i,0)}\right]^h W(y_i,\theta)\,dy_i=1. \tag{15.17}$$

Here, $W(y_i, \theta)$ -- unidimensional probability density of sample value y_i when signal strength equals θ. Therefore, $W(y_i, 0)$ and $W(y_i, \theta_1)$ -- unidimensional sample y_i distributions when there is no signal ($\theta = 0$) and when a signal with strength θ_1 is present, respectively.

Mean sample size $\overline{n(\theta)}$ is determined from the formula

$$\overline{n(0)}=\frac{L(\theta)\ln\beta_1+[1-L(\theta)]\ln\beta_2}{\overline{z(\theta)}}, \tag{15.18}$$

where

$$z(y_i) = \ln \frac{W(y_i, \theta_1)}{W(y_i, 0)}, \tag{15.19}$$

$$\overline{z(\theta)} = \int_{-\infty}^{\infty} z(y_i) W(y_i, \theta)\, dy_i, \tag{15.20a}$$

i. e., /268

$$\overline{z(\theta)} = \int_{-\infty}^{\infty} \left[\ln \frac{W(y_i, \theta_1)}{W(y_i, 0)}\right] W(y_i, \theta)\, dy_i. \tag{15.20b}$$

It follows from (15.15) and (15.18) that

$$\overline{n(0)} = \frac{(1 - P_{л т}) \ln \beta_1 + P_{л т} \ln \beta_2}{\overline{z(0)}} \tag{15.21}$$

and

$$\overline{n(\theta_1)} = \frac{P_{np}(\theta_1) \ln \beta_1 + [1 - P_{np}(\theta_1)] \ln \beta_2}{\overline{z(0)}}. \tag{15.22}$$

It follows from (15.13), (15.21), and (15.22) that, when $P_{л т} \leqslant 0{,}1$ and $P_{np}(\theta_1) \leqslant 0{,}1$, it is possible to assume

$$\left.\begin{aligned} \overline{n(0)} &\approx \frac{\ln \frac{1}{P_{np}(\theta_1)}}{-\overline{z(0)}} \\ &\text{and} \\ \overline{n(\theta_1)} &\approx \frac{\ln \frac{1}{P_{л т}}}{-\overline{z(\theta_1)}}. \end{aligned}\right\} \tag{15.23}$$

The following approximate formulas [16 and 85] are valid for magnitudes n(0) and n(θ_1) of variances $\sigma^2_{n(0)}$ and $\sigma^2_{n(\theta_1)}$, given slight $P_{л т}$ and $P_{np}(\theta_1)$:

$$\sigma^2_{n(0)} \approx \frac{\sigma^2_{z(0)} \ln \beta_1}{[\overline{z(0)}]^3} \tag{15.24}$$

and

$$\sigma^2_{n(\theta_1)} \approx \frac{\sigma^2_{z(\theta_1)} \ln \beta_2}{[\overline{z(\theta_1)}]^3}, \tag{15.25}$$

where

$$\sigma^2_{z(\theta)} = \overline{[z(\theta)]^2} - \left[\overline{z(\theta)}\right]^2; \tag{15.26}$$

$$\overline{[z(0)]^2} = \int_{-\infty}^{\infty} z^2(y_i)\, W(y_i, 0)\, dy_i; \tag{15.27}$$

$\sigma^2_{z(0)}$ and $\sigma^2_{z(\theta_1)}$ — variance $\sigma^2_{z(\theta)}$ values when $\theta = 0$ and $0 = \theta_1$, respectively.

It is possible to use formulas (15.16)--(15.26) to compute the main characteristics of the optimum sequential detector.

It follows from (15.16) that, when h = 0, magnitude L(θ) will become indeterminate. Magnitude $\overline{n(\theta)}$, computed from formula (15.18) accordingly /269 also turns out to be indeterminate. The following relationships are obtained after these ambiguities are expanded:

$$L(\theta') = \frac{\ln \beta_2}{\ln \dfrac{\beta_2}{\beta_1}} \tag{15.28}$$

and

$$\overline{n(\theta')} = \frac{\ln \beta_1 \ln \beta_2}{2\left[\dfrac{d\overline{z(\theta)}}{d\theta}\right]_{\theta=\theta'}} \left[\frac{dh}{d\theta}\right]_{\theta=\theta'}, \tag{15.29}$$

where θ' is value θ at which h = 0.

The next section will present an example illustrating use of the above-derived formulas and providing a representation of sequential detector quality.

15.4 Sequential Detection of a Fluctuating Pulse Signal

Let the signal comprise a train (packet) of independently-fluctuating pulses

and pass (along with noise) through an optimum linear filter matched with the signal, i. e., having bandwidth

$$\Pi \approx \frac{1}{\tau_u}, \qquad (15.30)$$

where τ_u — duration of each pulse.

An inertia-free linear detector, which separates the signal-plus-noise (or noise only) envelope, is connected at filter output. The distribution over time of the anticipated signal pulses is known at the point of reception and the task is to determine whether or not such a signal is present at receiver input.

Sequential analysis of voltage y(t) at detector output is used to solve this problem. Samples $y_1, y_2, \ldots, y_i$ of this voltage are taken at moments $t_1, t_2, \ldots, t_i$, corresponding to the moments the anticipated signal pulses cease at receiver input (as a signal pulse appears at optimum linear filter input, signal amplitude at filter output reaches its maximum at the moment the input pulse ceases).

Consequently, sample interval Δt equals

$$\Delta t = T_u,$$

where T_u — signal pulse spacing.

Since $T_u > \tau_u$, and filter band Π satisfies relationship (15.30), it is possible to consider that, given such an interval Δt selection, sampling values y_i mutually are statistically independent, both for noise alone (when there is no signal) and when a signal is present.

When there is no signal and only normal noise is active at detector /270
input, the distribution of y_i voltage values at linear detector output is subordinate to Rayleigh's law:

$$\left.\begin{array}{ll} W(y_l, 0) = \frac{y_l}{U_m^2} e^{-y_l^2/2U_m^2} & \text{where} \quad y_l > 0, \\ W(y_l, 0) = 0 & \text{where} \quad y_l < 0, \end{array}\right\} \qquad (15.31)$$

where U_m -- extant noise voltage value at detector input (i. e., at optimum filter output).

When the aforementioned signal is present, the distribution takes the form

$$\left.\begin{array}{ll} W(y_l, \mu^2) = \frac{y_l}{U_m^2 \mu^2} e^{-\frac{y_l^2}{2U_m^2\mu^2}} & \text{where} \quad y_l \geqslant 0; \\ W(y_l, \mu^2) = 0 & \text{where} \quad y_l < 0, \end{array}\right\} \qquad (15.32)$$

where, in accordance with (9.77) and (9.78)

$$\mu^2 = 1 + \frac{\overline{u_c^2}}{U_m^2} = 1 + \frac{Q_{cp}}{N_0 n}; \qquad (15.33)$$

Q_{cp}/n -- energy of one packet pulse; Q_{cp} -- energy of the entire packet (of n pulses).

We will designate:

$$\theta = \frac{Q_{cp}}{N_0 n}; \qquad (15.34)$$

then, $\theta = 0$, when there is no signal, while when there is a signal with anticipated average energy Q_{cp1}

$$\theta = \theta_1 = \frac{Q_{cp1}}{N_0 n}. \qquad (15.35)$$

Therefore, it is possible to write distribution (15.32) in the following form:

$$\left.\begin{array}{ll} W(y_l, \theta) = \frac{y_l}{U_m^2(1+\theta)} e^{-\frac{y_l^2}{2U_m^2(1+\theta)}} & \text{where} \quad y_l \geqslant 0, \\ W(y_l, \theta) = 0 & \text{where} \quad y_l < 0. \end{array}\right\} \qquad (15.36)$$

Substituting expressions (15.31) and (15.36) into formula (15.19), we have

$$z(y_l) = \ln\frac{1}{1+\theta_1} + \frac{\theta_1 y_l^2}{(1+\theta_1)\,2U_m^2}. \qquad (15.37)$$

It follows from (15.20a) and (15.37) that

$$\overline{z(\theta)} = \ln\frac{1}{1+\theta_1} + \frac{\theta_1 \overline{y_l^2}}{(1+\theta_1)\,2U_m^2},$$

where /271

$$\overline{y_l^2} = \int_0^\infty y_l^2 \frac{y_l}{U_m^2(1+\theta)} e^{-\frac{y_l^2}{2U_m^2(1+\theta)}} dy_l = 2U_m^2(1+\theta);$$

therefore

$$\overline{z(\theta)} = \ln\frac{1}{1+\theta_1} + \frac{\theta_1}{1+\theta_1}(1+\theta). \qquad (15.38)$$

Further, from (15.27), (15.36), and (15.37), we will find that

$$\overline{[z(\theta)]^2} = \left(\ln\frac{1}{1+\theta_1}\right)^2 + 2\left(\ln\frac{1}{1+\theta_1}\right)\frac{\theta_1}{(1+\theta_1)\,2U_m^2}\overline{y_l^2} + \left[\frac{\theta_1}{(1+\theta_1)\,2U_m^2}\right]^2 \overline{y_l^4}, \qquad (15.39)$$

where

$$\overline{y_l^4} = \int_0^\infty y_l^4 W(y_l, \theta)\, dy_l = 2(1+\theta)^2 4U_m^4. \qquad (15.40)$$

Substituting the expressions for $\overline{y_l^2}$ and $\overline{y_l^4}$ into (15.39), we obtain

$$\overline{[z(\theta)]^2} = \left(\ln\frac{1}{1+\theta_1}\right)^2 + 2\left(\ln\frac{1}{1+\theta_1}\right)\frac{\theta_1(1+\theta)}{(1+\theta_1)} + 2\left[\frac{\theta_1(1+\theta)}{1+\theta_1}\right]^2. \qquad (15.41)$$

From (15.38) and (15.41) we will find

$$\sigma^2_{z(\theta)} = \overline{[z(\theta)]^2} - [\overline{z(\theta)}]^2 = \left[\frac{\theta_1(1+\theta)}{(1+\theta_1)}\right]^2. \qquad (15.42)$$

Hence, it follows that

$$\sigma^2_{z\,(0)} = \left(\frac{\theta_1}{1+\theta_1}\right)^2 \quad \text{and} \quad \sigma^2_{z\,(\theta_1)} = \theta_1^2. \tag{15.43}$$

From (15.21), (15.22), and (15.38) we have

$$\left.\begin{aligned} \overline{n\,(0)} &= \frac{(1-P_{\text{лт}})\ln\beta_1 + P_{\text{лт}}\ln\beta_2}{\dfrac{\theta_1}{1+\theta_1} - \ln(1+\theta_1)}; \\ \overline{n\,(\theta_1)} &= \frac{P_{\text{пр}}(\theta_1)\ln\beta_1 + [1-P_{\text{пр}}(\theta_1)]\ln\beta_2}{\theta_1 - \ln(1+\theta_1)}. \end{aligned}\right\} \tag{15.44}$$

From formulas (15.24) and (15.25) we will find [considering relationships (15.38) and (15.43)] that

$$\left.\begin{aligned} \sigma^2_{n\,(0)} &= \frac{\left(\dfrac{\theta_1}{1+\theta_1}\right)^2 \ln\beta_1}{\left[\dfrac{\theta_1}{1+\theta_1} - \ln(1+\theta_1)\right]^3}; \\ \sigma^2_{n\,(\theta_1)} &= \frac{\theta_1^2 \ln\beta_2}{[\theta_1 - \ln(1+\theta_1)]^3}. \end{aligned}\right\} \tag{15.45}$$

It is not difficult using formulas (15.44), (15.45), and (15.13) to /272
compute the average sizes of samples $\overline{n(0)}$ and $\overline{n(\theta_1)}$ and their variances $\sigma^2_{n\,(0)}$ and $\sigma^2_{n\,(\theta_1)}$ for given values θ_1, $P_{\text{пр}}(\theta_1)$ and $P_{\text{лт}}$.

For example, let $\theta_1 = 1$; $P_{\text{пр}}(\theta_1) = 0.1$ and $P_{\text{лт}} = 10^{-3}$. Then, from the aforementioned formulas, we will obtain

$$\left.\begin{aligned} \overline{n\,(0)} &= 11{,}5; \quad \overline{n\,(\theta_1)} = 20. \\ \sigma_{n\,(0)} &= 8{,}5; \quad \sigma_{n\,(\theta_1)} = 16. \end{aligned}\right\} \tag{15.46}$$

Here

$$\frac{\sigma_{n\,(0)}}{\overline{n\,(0)}} = 0{,}74 \quad \text{and} \quad \frac{\sigma_{n\,(\theta_1)}}{\overline{n\,(\theta_1)}} = 0{,}8.$$

It is significantly more difficult to compute characteristics $L(\theta)$ and $\overline{n(\theta)}$ for <u>random</u> signal strength values θ (and not only for $\theta = 0$ and $\theta = \theta_1$).

To do so requires that supplemental parameter $h = h(\theta)$ first be found from functional equation (15.17).

In the case under examination where distribution $W(y_i, \theta)$ is described by expression (15.36), equation (15.17) takes the form

$$\int_0^\infty \left[\frac{1}{1+\theta_1} e^{\frac{\theta_1 y_i^2}{(1+\theta_1)\, 2U_m^2}}\right]^h \frac{y_i^2}{U_m^2(1+\theta)} e^{\frac{-y_i^2}{(1+\theta)\, 2U_m^2}} dy_i = 1.$$

After transformation [85], this equation takes the following form:

$$(1+\theta_1)^h - h\theta_1(1+\theta_1)^{h-1}(1+\theta) = 1. \tag{15.47}$$

It is transcendental relative to h. Therefore, one must find the dependence of h on θ by solving equation (15.47) for θ, rather than h:

$$\theta = \frac{(1+\theta_1) - (1+\theta_1)^{1-h}}{\theta_1 h} - 1. \tag{15.48}$$

It is not difficult using this formula for a given θ_1 to plot the dependence of θ on h. This dependence is depicted in Figure 15.7 for $\theta_1 = 1$. Here,

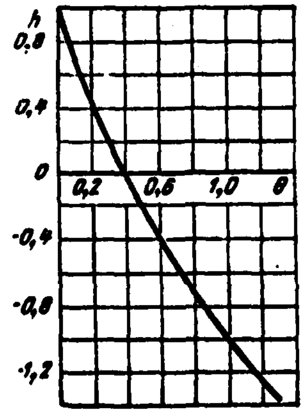


Figure 15.7

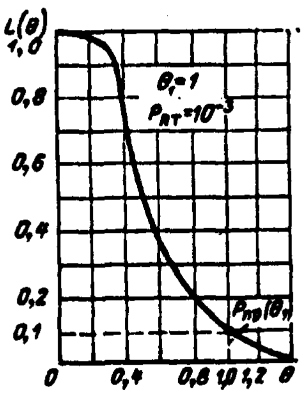


Figure 15.8

the following corresponds to value h = 0

$$\theta = \theta' = 0{,}4. \tag{15.49}$$

It is possible to use the Figure 15.7 curve to compute dependence $L(\theta)$ determined from formula (15.16). Here, at the ambiguity point ($\theta = \theta'$, h = 0), formula (15.28) rather than formula (15.16) should be used for the computation.

Characteristic $L(\theta)$ depicted in Figure 15.8 and corresponding to $P_{\text{пр}}(\theta_1) = 0.1$, $P_{\text{лт}} = 10^{-3}$, and $\theta_1 = 1$ is computed in the same way.

At points $\theta = 0$ and $\theta = \theta_1 = 1$ this characteristic, as would be expected, provides these initial magnitudes:

$$L(0) = 1 - P_{\text{лт}} = 0{,}999,$$

$$L(\theta_1) = P_{\text{пр}}(\theta_1) = 0{,}1.$$

/273

Dependence $\overline{n(\theta)}$ is computed from formula (15.18). Here, values $L(\theta)$ are taken from the Figure 15.8 graph, while magnitude $\overline{z(\theta)}$ is computed from formula (15.38).

At the ambiguity point ($\theta = \theta'$, h = 0), magnitude $\overline{n(\theta')}$ should be computed from formula (15.29).

In the case under examination, from (15.38) we have

$$\left[\frac{d\overline{z(\theta)}}{d\theta}\right]_{\theta=\theta'} = \frac{\theta_1}{1+\theta_1}. \tag{15.50}$$

From (15.48) we will find

$$\left(\frac{d\theta}{dh}\right)_{h=0} = -\frac{(1+\theta_1)\,[\ln(1+\theta_1)]^2}{2\theta_1};$$

therefore

$$\left(\frac{dh}{d\theta}\right)_{\theta=\theta'} = -\frac{2\theta_1}{(1+\theta_1)\,[\ln(1+\theta_1)]^2}\,. \tag{15.51}$$

Substituting (15.50) and (15.51) into (15.29), we obtain

$$\overline{n(\theta')} = \frac{\lg \beta_2 \lg \frac{1}{\beta_1}}{[\lg(1+\theta_1)]^2}\,. \tag{15.52}$$

In the particular case being examined, when $\theta_1 = 1$, $P_{пр}(\theta_1) = 0.1$, and $P_{лт} = 10^{-3}$, this formula provides

$$\overline{n(\theta')} = 31.$$

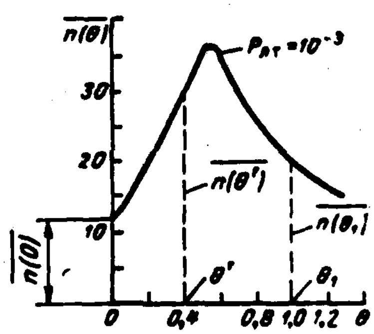


Figure 15.9

Figure 15.9 depicts characteristic $\overline{n(\theta)}$ plotted in this manner. Where /274 $\theta = 0$ and $\theta = \theta_1$, it has the values $\overline{n(\theta)} = 11.5$ and $\overline{n(\theta_1)} = 20$ already found above.

15.5 Comparison of Sequential Detection to Classic Detection

Now, we will compare results obtained during sequential detection to corresponding results from classic binary detection.

It is possible to characterize the gain obtained during sequential detection by the ratios:

$$\left.\begin{array}{c}\dfrac{n}{\overline{n(0)}},\ \dfrac{n}{\overline{n(\theta_1)}}\ \text{and}\ \dfrac{n}{\bar{n}},\\ \bar{n}=P(0)\overline{n(0)}+P(C)\overline{n(\theta_1)}.\end{array}\right\} \tag{15.53}$$

where

Here, n -- sample size during classic binary detection with identical signal strength values θ_1 and error probabilities $P_{лт}$ and $P_{пр}$.

The Figure 12.9 curves from classic binary detection are valid in this case of detection of a train of independently-fluctuating signal pulses. In this figure, ratio $Q_{ср}/N_0$, where $Q_{ср}$ -- total energy of n pulses, is plotted on the Y-axis. Therefore, in accordance with (15.34), we have

$$\frac{Q_{ср}}{N_0}=n\theta,$$

while, for anticipated signal strength θ_1

$$\frac{Q_{ср}}{N_0}=n\theta_1. \tag{15.54}$$

We will do a comparison for the following values of magnitudes θ_1, $P_{лт}$, and $P_{пр}(\theta_1)$:

$$\theta_1=1;\ P_{лт}=10^{-3};\ P_{пр}(\theta_1)=0{,}1,$$

i. e.,

$$P_{по}=1-P_{пр}(\theta_1)=0{,}9.$$

Here, formula (15.54) provides

$$\frac{Q_{ср}}{N_0}=n,$$

and, from the Figure 12.9 curve corresponding to $P_{лт} = 10^{-3}$ and $P_{по} = 0.9$, we will find that

$$n = 35. \tag{15.55}$$

Relationships (15.46) were obtained in §15.4 for sequential analysis /275 of the same $P_{лт}$ and $P_{пр}(\theta_1)$ values. Therefore, from (15.46), (15.53), and (15.55), we have

$$\frac{n}{\overline{n(0)}} = 3{,}05; \quad \frac{n}{\overline{n(\theta_1)}} = 1{,}75;$$
$$\frac{n}{\bar{n}} = \frac{35}{11{,}5P(0) + 20[1 - P(0)]} = \frac{1{,}75}{1 - 0{,}42P(0)}.$$

It follows from the last relationship that gain $n/\bar{n}$ in an average (unconditional) sample size will depend on a priori signal absence probability P(0) and increases when this probability rises. The following result when probability P(0) equals 0, 0.5, and 1, respectively:

$$\frac{n}{\bar{n}} = 1{,}75; \quad \frac{n}{\bar{n}} = 2{,}2 \quad \text{and} \quad \frac{n}{\bar{n}} = 3{,}05.$$

Hence, it follows that, in the case under examination, the gain in an average sample size (i. e., in mean observation duration $\bar{T}$) is 1.75 ÷ 3.05 [depending on a priori probability P(0)]. Here, the greatest gain magnitude (3.05) corresponds to value P(0) = 1 or [where P(0) ≠ 1] to cases when there is no signal.

As computations demonstrate [85], given a decrease in permissible false-alarm probability and all other things being equal, gain $n/\overline{n(0)}$ monotonically increases and may significantly exceed unity. For example, where $\theta_1 = 1$, $P_{пр}(\theta_1) = 0.9$, and $P_{лт} = 10^{-6}$, the result is

$$\frac{n}{\overline{n(0)}} = 5{,}2.$$

The aforementioned examples corresponded to a case where $P_{лт} \ll P_{пр}(\theta_1)$. If, in accordance with problem conditions,

$$P_{\text{лт}} = P_{\text{пр}}(\theta_1),$$

then, as computations for several signal types made by different authors show, the result is

$$\frac{n}{\overline{n(0)}} \approx \frac{n}{\overline{n(\theta_1)}} \approx 2.$$

Since the result here is

$$\overline{n(0)} \approx \overline{n(\theta_1)},$$

then, from (15.53), we obtain

$$\overline{n} = \overline{n(0)}\,[P(0) + P(C)] = \overline{n(0)}$$

and, consequently,

$$\frac{n}{\overline{n}} \approx 2.$$

It follows from these data that, when $P_{\text{лт}} \approx P_{\text{пр}}(\theta_1)$, sequential analysis /276
provides a gain in mean observation duration by approximately a factor of 2. If $P_{\text{лт}} \ll P_{\text{пр}}(\theta_1)$, then the gain may be significantly greater, especially if signal absence probability P(0) is sufficiently high or if the gain is assessed only by ratio $n/\overline{n(0)}$, i. e., only for a case of no signal. Therefore, it is most advisible to use sequential analysis when problem conditions stipulate that $P_{\text{лт}} \ll P_{\text{пр}}(\theta_1)$ and when the a priori signal absence probability is significantly greater than the a priori signal present probability, i. e., $P(0) \gg P(C)$. This situation arises often in radar.

However, along with the aforementioned advantage, sequential analysis has major drawbacks compared to classic analysis.

The first drawback is the randomness of sample size n, i. e., of observation duration T.

In the example examined in § 15.4 [see relationships (15.46)], sample size standard deviation σ_n from its mean value is 74--80%. Computations made for several signal types by different authors demonstrate that standard deviation σ_n from the average sample size is 50--80% for other typical signals as well. This signifies that, in a rather-significant percent of sequences, observation time may exceed its mean value somewhat.

So-called split sequential analysis is used to ameliorate this shortcoming. Here, a predetermined maximum observation time value $T_{макс}$ is established and analysis mandatorily must cease when this time runs out (if it has not already ended).

As long as observation duration does not exceed $T_{макс}$, i. e., the number of steps i does not exceed magnitude

$$n_{макс}=\frac{T_{макс}}{\Delta t},$$

sequential observation will be conducted with respect to the steps in the normal manner, i. e., with two thresholds β_1 and β_2. In step $n_{макс}$, only one threshold β_3 instead of two thresholds β_1 and β_2 is established (as in classic binary detection). Therefore, if observation has not yet concluded prior to transition to this step, then it will conclude mandatorily at this step by making the decision "no" (if it turns out that $l_{n\,макс}<\beta_3$) or the decision "yes" (if $l_{n\,макс}>\beta_3$).

Splitting the observation into steps $n_{макс}$ rules out the possibility of very-long observation durations T significantly exceeding $\bar{T}$. However, the lower the $T_{макс}$ selected, i. e., the more significant the splitting, the less the gain $T/\bar{T}$ in mean observation time obtained during sequential analysis. This is explained by the fact that, when ratio $T_{макс}/\bar{T}$ decreases, sequential analysis becomes /277 more like classic analysis.

The second shortcoming of sequential analysis compared to classic analysis

is its great complexity. Sequential analysis may be simplified considerably by preliminary conversion of analog sampling values y_i into quantum binary magnitudes --ones and zeros.

This simplified method is examined in the following section.

15.6 Sequential Analysis of Quantum Samples

In the sequential analysis approach under examination, sample values y_1, y_2, . . ., y_i beforehand are converted into ones and zeros for some quantization

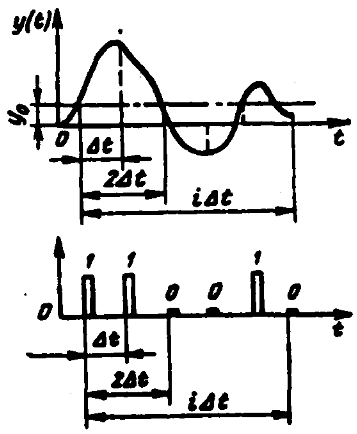


Figure 15.10

threshold y_0 (Figure 15.10): if $y_i > y_0$, then a one is produced at quantizer output; otherwise, a zero is produced (where $y_i < y_0$).

Thus, ones and zeros, which are subjected to sequential analysis, are formed at quantizer output. Interval Δt is such that all observation results (i. e., all ones and zeros) among themselves are statistically independent.

It usually is assumed for the purposes of further simplification that usable information (i. e., information as to whether a signal is present or not) will be contained exclusively in the quantity of ones (or zeros) in a given observation

interval ($i\Delta t$), i. e., disregarding the influence of the position of the ones over time within this interval.

The probability that there will be k_i ones and, consequently, $(i-k_i)$ zeros in interval $i\Delta t$ is determined by known binomial distribution

$$W_\theta(k_i,\ y_0)=\frac{i!}{k_i!\,(i-k_i)!}P_\theta^{k_i}(1-P_\theta)^{i-k_i} \qquad (k_i=0,\ 1,\ 2,\dots,\ i), \tag{15.56}$$

where

$$P_\theta=P_\theta(y_i>y_0)$$

is the probability that y_i will exceed y_0, i. e., the probability of the appearance of ones at quantizer output.

Index θ in magnitudes $W_\theta(k_i, y_0)$ and P_θ underscores the circumstance /278 that these magnitudes will depend on signal strength θ. Therefore, instead of (15.56), it is possible to write

$$W_{\theta_1}(k_i,\ y_0)=\frac{i!}{k_i!\,(i-k_i)!}P_{\theta_1}^{k_i}(1-P_{\theta_1})^{i-k_i}, \tag{15.57a}$$

$$W_0(k_i,\ y_0)=\frac{i!}{k_i!\,(i-k_i)!}P_0^{k_i}(1-P_0)^{i-k_i}, \tag{15.57b}$$

where P_{θ_1} and P_0 -- probabilities of the appearance of ones given presence of a signal with anticipated strength θ_1 and given absence of a signal ($\theta = 0$), respectively, while $W_{\theta_1}(k_i, y_0)$ and $W_0(k_i, y_0)$ -- binomial distributions for presence of a signal of intensity θ_1 and when there is no signal ($\theta = 0$), respectively.

It follows from (15.57) that the logarithm of the likelihood factor in this case equals [also see (15.19)]

$$z(i,\ y_0)=\ln\frac{W_{\theta_1}(k_i,\ y_0)}{W_0(k_i,\ y_0)}=\ln\frac{P_{\theta_1}^{k_i}(1-P_{\theta_1})^{i-k_i}}{P_0^{k_i}(1-P_0)^{i-k_i}},$$

or

$$z(i,\ y_0)=k_i\left(\ln\frac{P_{\theta_1}}{P_0}-\ln\frac{1-P_{\theta_1}}{1-P_0}\right)+i\ln\frac{1-P_{\theta_1}}{1-P_0}. \tag{15.58}$$

In accordance with the rules of sequential analysis, magnitude $z(i,\ y_0)$ in each step i must be compared with thresholds β_1 and β_2, determined from formulas (15.13):

$$\left.\begin{array}{l}\text{if } z(i,\ y_0)<\beta_1\text{, then hypothesis } H_0 \text{ (no signal) is accepted;}\\ \text{if } z(i,\ y_0)>\beta_2\text{, then hypothesis } H_1 \text{ (signal with strength } \theta_1\text{)}\\ \text{is accepted;}\end{array}\right\} \tag{15.59}$$

finally, if it turns out that

$$\beta_1\leqslant z(i,\ y_0)\leqslant\beta_2, \tag{15.60}$$

then the test continues, i. e., there is a transition to the next observation step.

It is simple to obtain the following decision areas based on relationships (15.58)--(15.60):

$$\left.\begin{array}{l}\text{if } k_i<\frac{b+i}{c+1}\text{, then hypothesis } H_0 \text{ is accepted;}\\ \text{if } k_i>\frac{a+i}{c+1}\text{, then hypothesis } H_1 \text{ is accepted;}\\ \text{if } \frac{b+i}{c+1}\leqslant k_i\leqslant\frac{a+i}{c+1}\text{, then the test (observation) continues.}\end{array}\right\} \tag{15.61}$$

Here /279

$$a=\frac{\lg\beta_2}{\lg\frac{1-P_0}{1-P_{\theta_1}}};\quad b=\frac{\lg\beta_1}{\lg\frac{1-P_0}{1-P_{\theta_1}}};\quad c=\frac{\lg\frac{P_{\theta_1}}{P_0}}{\lg\frac{1-P_0}{1-P_{\theta_1}}}. \tag{15.62}$$

It is possible to draw the following conclusions from examination of relationships (15.61) and (15.62):

1. A sequential detector completely is determined by three parameters a, b, and c, which are functions of four parameters: β_1, β_2, P_0, and P_{θ_1}.

2. Parameters a, b, and c will depend, among other things, on quantization threshold y_0 (since probabilities P_{θ_1} and P_0 will depend on this threshold). Therefore, an appropriate (optimum) quantization threshold must be selected to insure better sequential detector operation.

3. Since parameters a, b, and c remain unchanged during the sequential analysis process, then this analysis will boil down to computation of the number of ones k_i at quantizer output and comparison of this number with thresholds in accordance with very-simple relationships (15.61).

It follows from what has been stated that the sequential detector being examined is comparatively simple in design. Its basic elements are a quantizer, which converts observed signal-plus-noise y(t) into a sequence of ones and zeros, and a binary scaler.

Introduction of quantization will lead to some loss in usable information (see Figure 15.10). Therefore, sequential analysis of quantized samples in principle must provide worse results than optimum sequential analysis of non-quantized sample values described in preceding sections. However, as theoretical and experimental analyses done by several authors [16, 85, and 121] have shown, the resultant loss in typical cases is relatively slight. Thus, for example, during detection of a sinusoidal carrier (and optimum selection of quantization threshold y_0), introduction of quantization will lead to a mean observation time increase by a factor of approximately 1.5 [16].

Therefore, in a number of cases, significant detector design simplification, which accompanies introduction of quantization, completely may compensate for the resultant slight deterioration in detection quality.

CHAPTER SIXTEEN /280

ANALOG MESSAGE RECEPTION ANALYSIS USING THE DISTRIBUTION PARAMETER ESTIMATE APPROACH

16.1 Problem Formulation in the Classic Theory of Statistical Estimates

Initially, we will examine the basic assumptions of the so-called classic theory of statistical estimates, i. e., a theory formulated several decades ago, and then we will explain how this theory applies to solving the problem of the reception of analog messages on a noise background. Use of the latest theory of statistical estimates, the foundation for which was laid in the works of A. Wald [30], is examined in Chapters 17 and 18.

In the classic theory of estimates, the problem is formulated in the following manner:

A sample of n random magnitudes $y_1, y_2, \ldots, y_n$ exists (they need not be statistically independent) with a precisely-known law of distribution, except for some parameters $\alpha_1, \alpha_2, \ldots, \alpha_m$. For simplicity, initially we will formulate the problem relative to the case when only one distribution α parameter is unknown. This signifies that, for a given α, the following n-dimensional probability density is known

$$P_\alpha(y) = L(y_1, \ldots, y_n | \alpha). \tag{16.1}$$

The task is to establish the magnitude of parameter α based on the given sample $(y_1, \ldots, y_n)$.

The assumption is that, in the process of the given sample, parameter α remains unchanged. During transition from one sequence to the next (i. e., from one sample to another), desired parameter α may change like a random magnitude or remain unchanged.

Since sample values $y_1, \ldots, y_n$ are random magnitudes, and sample size n is finite, then it is impossible to establish the parameter α magnitude with sufficient accuracy based on sample analysis. It is possible only to draw an approximate conclusion as to its magnitude or, as they say, to make an estimate. Here, the following approaches, which differ in principle, to estimating the needed parameter are possible:

1. Point estimate.

2. Interval estimate.

In a point estimate, some value α^*, being a function of sample values $(y_1, \ldots, y_n)$ and accepted as the true parameter value, is supplied as a result of sample analysis. This value α^* is called the parameter α estimate.

Function type

$$\alpha^* = f(y_1, \ldots, y_n) \tag{16.2}$$

is selected so that estimate α^* is as close as possible (in some predetermined /281 sense) to true value α, i. e., so the error

$$\delta = \alpha^* - \alpha, \tag{16.3}$$

arising during the estimate is as small as possible.

As a result of the randomness of α^* and α, error δ also is a random magnitude, which changes from one sequence to another. Therefore, the complete

composite law of distribution of magnitudes α^* and α characterizes the quality of the estimate

$$P(\alpha^*, \alpha) = P(\alpha) P_\alpha(\alpha^*). \tag{16.4}$$

However, in many cases, a priori distribution $P(\alpha)$ of needed parameter α is unknown, while, in several cases, parameter α is in no way random due to its physical nature. In these cases, distribution $P_\alpha(\alpha^*)$ of estimate α^* for a given parameter α value must suffice instead of unconditional distribution $P(\alpha^*, \alpha)$ (this law must be found for all possible values of α). But, in several cases, it is very difficult or even practically impossible to find law of distribution $P_\alpha(\alpha^*)$.

Consequently, in such cases, the quality (validity) of the estimate α^* found turns out to be fully or partially known. This circumstance is the main drawback inherent in a point estimate. However, as will be explained subsequently, given a good-enough estimate approach [i. e., given sufficiently successful selection of the type of function (16.2)] and large sample size (n $\gg 1$), law of distribution $P_\alpha(\alpha^*)$ turns out to be normal or close to normal and, consequently, error δ sufficiently is characterized by its mean value

$$\overline{\delta}_\alpha = M_\alpha(\alpha^* - \alpha) = M_\alpha \alpha^* - \alpha \tag{16.5}$$

and mean square

$$\overline{\delta_\alpha^2} = M_\alpha(\alpha^* - \alpha)^2. \tag{16.6}$$

Here, index α denotes that averaging is conditional, i. e., done given that magnitude α is unchanged.

Magnitude $\overline{\delta}_\alpha$ also is called estimate bias. Evidently, the less bias $\overline{\delta_\alpha}$ and mean square $\overline{\delta_\alpha^2}$, the higher the quality of the estimate.

In many cases, it is considerably simpler to compute parameters $\overline{\delta_\alpha}$ and $\overline{\delta_\alpha^2}$ than it is to find law of distribution $P_\alpha(\alpha^*)$. Therefore, large samples

($n \gg 1$) and a good estimate method significantly ameliorate the aforementioned point estimate shortcoming.

But, given a correct estimate method, an increase sample n size provides an increase in estimate accuracy. Therefore, it also is possible to assert that the more accurate the estimate with respect to problem conditions, the more basis exists for use of a point estimate.

If, on the other hand, according to problem conditions, the sample size /282 is small and the estimate may not be precise enough, use of a point estimate usually turns out to be inadvisible since, here, due to the aforementioned shortcoming, one cannot obtain sufficiently-valid data concerning the quality of estimate α^* found. In such cases (given small samples), the interval estimate method usually is used in statistics. This approach is based on introduction of so-called confidence intervals.

In this approach, the result of the analysis of sample ($y_1, \ldots, y_n$) is not the specific magnitude α^* of the needed parameter (as is the case in a point estimate), but only interval $\alpha_1 \div \alpha_2$, which will comprise needed magnitude with a probability equalling some preselected magnitude ($1-\epsilon$), called the confidence coefficient. The latter is close to unity (0.99, for example), i. e., $\epsilon \ll 1$, in order to obtain sufficiently-high confidence in the estimate.

Interval $\alpha_1 \div \alpha_2$ is called the confidence interval corresponding to a given confidence coefficient.

This estimate mathematically boils down to the following. Two sample value functions $\alpha_1(y_1, \ldots, y_n)$ and $\alpha_2(y_1, \ldots, y_n)$ are selected so that the following relationship is satisfied for any value of

$$P[\alpha_1(y_1, \ldots, y_n) < \alpha < \alpha_2(y_1, \ldots, y_n)] = 1-\varepsilon. \qquad (16.7)$$

Here, as usual, P[ ] denotes the probability that inequalities enclosed in the square brackets are satisfied.

Since sample values $y_1, \ldots, y_n$ are random magnitudes, then limits α_1

and α_2 of the confidence interval also are random magnitudes changing from one test (one sample) to another.

Needed parameter α by nature does not have to be random and, in relationship (16.7), parameter α is considered simply some fixed magnitude irrespective of whether it is random or not. Therefore, one should treat the left portion of relationship (16.7) as the "probability that interval $\alpha_1 \div \alpha_2$, includes needed parameter α " rather than the "probability that α will be found within the limits of interval $\alpha_1 \div \alpha_2$ ".

The form of functions $\alpha_1(y_1, \ldots, y_n)$ and $\alpha_2(y_1, \ldots, y_n)$ should be such that, for a given confidence coefficient $(1-\varepsilon)$, interval $(\alpha_1 \div \alpha_1)$ will be as small as possible. This interval is a random magnitude changing from sample to sample. Therefore, we have no assurances that, in all tests, this interval will be small. But, a drawback of this type is inherent in a point estimate too, since point estimate error δ also in individual cases may not be slight.

The advantage of the interval estimate method is that, as a result of the estimate, i. e., after finding interval $\alpha_1 \div \alpha_2$, we obtain not only the estimate itself (i. e., an indication that needed parameter α will be found within the limits $\alpha_1 \div \alpha_2$ found), but also we know the quality of the resultant estimate beforehand--we know that the probability that needed parameter α is encompassed by resultant interval $\alpha_1 \div \alpha_2$ equals previously-established (and, consequently, /283
acceptable) magnitude $(1-\varepsilon)$. Therefore, the confidence interval approach is applicable for small samples as well.

It follows from the aforementioned point estimate and interval estimate properties that the point estimate is more convenient, given large samples ($n \gg 1$), which allow slight point estimate error $\bar{\delta}^2$. On the other hand, an interval estimate usually turns out to be more convenient for small samples.

Both estimate approaches also are used to solve problems of reception of signals on a noise background. However, the point estimate is used significantly more often than is the interval estimate. This is because, first, signal (message) reception usually requires very high message reproduction accuracy and resultant sample size n is so large that the main point estimate shortcoming mentioned above

barely manifests itself. Second, the result of the estimate when reproducing signals (or messages) usually must be provided automatically and continuously at receiver output. Here, naturally, it is less convenient (and sometimes simply unacceptable) to form in a receiver two functions $\alpha_1(y_1, \ldots, y_n)$ and $\alpha_2(y_1, \ldots, y_n)$ and to obtain interval $\alpha_1 \div \alpha_2$ than it is to form and provide magnitude $\alpha^*(y_1, \ldots, y_n)$ at output. Therefore, in future, we will limit ourselves to examination only of a point estimate. Here, for simplicity, we will discard the term "point" and will use, as is the convention in foreign literature, the single term "estimate" for the following concepts: "estimate process," "estimate method" [function $\alpha^*(y_1, \ldots, y_n)$ type], and "estimate result."

For example, if one says that α^* is an efficient estimate, then this denotes that the estimate process, estimate method [function $\alpha^*(y_1, \ldots, y_n)$ type], and estimate result are efficient.

16.2 Basic Relationships During a Point Estimate

If certain sufficiently-general differentiability conditions [23] are met, then an estimate is called <u>sufficient</u> and the mean square $\overline{\delta_\alpha^2}$ of the error arising during the estimate satisfies inequality

$$\overline{\delta_\alpha^2} \geqslant \frac{\left(\frac{\partial M_\alpha \alpha^*}{\partial \alpha}\right)^2}{M_\alpha\left(\frac{\partial \ln L}{\partial \alpha}\right)^2}, \tag{16.8}$$

where L -- function determined from relationship (16.1) and called the likelihood function.

If estimate α^* is such that there is an equal sign in relationship (16.8), then such an estimate is called <u>efficient</u>.

Consequently, for an efficient estimate, mean square error $\overline{\delta_\alpha^2}$ has the minimum possible value and this value is determined from the right side of inequality /284 (16.8). However, an efficient estimate exists when and only when likelihood function $L(y_1, \ldots, y_n | \alpha)$ satisfies two special conditions [23]. We must shift from random

variables $y_1, \ldots, y_n$ to random variables $\xi_1, \ldots, \xi_{n-1}, \alpha^*$ linked on a one-to-one basis to them in order to formulate these conditions.

Let $h(\xi_1, \ldots, \xi_{n-1}|\alpha^*, \alpha)$ be the composite probability density of variables $\xi_1, \ldots, \xi_{n-1}$ for givens α^* and α, while $h(\alpha^*|\alpha)$ -- probability density of estimate α^* for a given α value. Then, the aforementioned two conditions have the following form:

A. Function $h(\xi_1, \ldots, \xi_{n-1}|\alpha^*, \alpha)$ will not depend on α. (16.9)

B. This inequality is satisfied

$$\frac{\partial \ln h(\alpha^*|\alpha)}{\partial \alpha} = k(\alpha^* - \alpha), \tag{16.10}$$

where k does not depend on α^* (but may depend on α).

If only condition A of the above two conditions is satisfied, then estimate α^* is called sufficient.

It follows from the nature of conditions A and B that efficient and even sufficient estimates may exist in far from all the cases of practical interest to us. Therefore, in the general case, inequality (16.8) makes it possible to determine only the lower mean square error limit.

Along with mean square error $\overline{\delta_\alpha^2}$, bias magnitude $\overline{\delta_\alpha}$ [see (16.5)] also is an important characteristic of estimate α^* quality.

Estimate α^* having zero bias is called unbiased. Consequently, these relationships are valid for an unbiased estimate

$$\overline{\delta_\alpha} = 0, \quad \text{i. e.,} \quad M_\alpha \alpha^* = \alpha. \tag{16.11}$$

It follows from relationships (16.8) and (16.11) that inequality (16.8) is simplified for unbiased estimates and takes the form

$$\overline{\delta_\alpha^2} \geqslant \frac{1}{M_\alpha\left(\frac{\partial \ln L}{\partial \alpha}\right)^2}. \tag{16.12}$$

When sample size n increases, estimate quality improves; therefore, when $n \to \infty$, estimate α^* may in some cases acquire those valuable properties (efficiency and/or nonbias), which it lacked when n was small.

An estimate, which will become unbiased when $n \to \infty$, is called consistent, while an estimate, which becomes efficient when $n \to \infty$, is called asymptotically efficient.

Finally, in many cases, the estimate α^* law of distribution will become normal when $n \to \infty$. Such α^* estimates are called asymptotically normal.

The aforementioned determinations may be formulated in the following way:

Estimate α^* is consistent if inequality (16.11) is satisfied when $n \to \infty$. /285

Estimate α^* is asymptotically efficient if mean-square error $\overline{\delta_\alpha^2}$ attains the minimum-possible value determined by the right side of inequality (16.8) when $n \to \infty$.

If sample values $y_1, \ldots, y_n$ statistically are independent, then

$$L(y_1, \ldots, y_n|\alpha) = f(y_1|\alpha) f(y_2|\alpha) \ldots f(y_n|\alpha), \tag{16.13}$$

where $f(y_i|\alpha)$ -- unidimensional probability density of magnitude y_i for a given α.

Here, inequality (16.8) takes the form

$$\overline{\delta_\alpha^2} \geqslant \frac{\left(\frac{\partial M_\alpha \alpha^*}{\partial \alpha}\right)^2}{nM_\alpha\left[\frac{\partial \ln f(y_i \mid \alpha)}{\partial \alpha}\right]^2}. \tag{16.14}$$

Magnitude $M_\alpha\left[\frac{\partial \ln f(y_i \mid \alpha)}{\partial \alpha}\right]^2$ also may be represented in the form of the following expressions:

$$M_\alpha\left[\frac{\partial \ln f(y_i \mid \alpha)}{\partial \alpha}\right]^2 = \int_{-\infty}^{\infty} f(y_i \mid \alpha)\left[\frac{\partial \ln f(y_i \mid \alpha)}{\partial \alpha}\right]^2 dy_i = \\ = \int_{-\infty}^{\infty} \frac{1}{f(y_i \mid \alpha)}\left[\frac{\partial f(y_i \mid \alpha)}{\partial \alpha}\right]^2 dy_i. \tag{16.15}$$

All the aforementioned relationships applied to a case when the distribution of samples (y_1, . . ., y_n) will comprise only one unknown parameter α.

Now, we will examine a case when this distribution has two unknowns, α_1 and α_2. In this case, the likelihood function has the form

$$P_{\alpha_1, \alpha_2}(y) = L(y_1, \ldots, y_n \mid \alpha_1, \alpha_2). \tag{16.16}$$

We will assume that both unknowns will be subject to estimation and we will designate the results of the estimate α_1^* and α_2^*.

Estimates α_1^* and α_2^* are sample value functions, i. e.,

$$\alpha_1^* = \alpha_1^*(y_1, \ldots, y_n) \text{ and } \alpha_2^* = \alpha_2^*(y_1, \ldots, y_n). \tag{16.17}$$

For notational simplicity, we will limit ourselves to a case of unbiased estimates. Here,

$$M_{\alpha_1} \alpha_1^* = \alpha_1; \tag{16.18}$$

$$M_{\alpha_2} \alpha_2^* = \alpha_2 \tag{16.19}$$

and, consequently, /286

$$\left.\begin{aligned}\overline{\delta^2_{\alpha_1}} = M_{\alpha_1}(\alpha_1^* - \alpha_1)^2 = M_{\alpha_1}(\alpha_1^* - M_{\alpha_1}\alpha_1^*)^2 = \sigma^2_{\alpha_1}\\ \overline{\delta^2_{\alpha_2}} = M_{\alpha_2}(\alpha_2^* - \alpha_2)^2 = M_{\alpha_2}(\alpha_2^* - M_{\alpha_2}\alpha_2^*)^2 = \sigma^2_{\alpha_2},\end{aligned}\right\} \tag{16.20}$$

i. e., in a case of unbiased estimates, mean square errors $\overline{\delta^2_{\alpha_1}}$ and $\overline{\delta^2_{\alpha_2}}$ coincide with variances $\sigma^2_{\alpha_1}$ and $\sigma^2_{\alpha_2}$ of these errors (here, both mean squares and variances are conditional, i. e., found through averaging for given α_1 and α_2 values).

During estimation of two parameters α_1 and α_2, besides error $\sigma^2_{\alpha_1}$ and $\sigma^2_{\alpha_2}$ variances, often it is important to know mutual correlation factor ρ_{α_1,α_2} of these errors, which for unbiased estimates by definition equals

$$\rho_{\alpha_1,\alpha_2} = \frac{M_{\alpha_1,\alpha_2}(\alpha_1^* - \alpha_1)(\alpha_2^* - \alpha_2)}{\sigma_{\alpha_1}\sigma_{\alpha_2}}. \tag{16.21}$$

Based on parameters σ_{α_1}, σ_{α_2} and ρ_{α_1,α_2}, it is possible to plot a dispersion ellipse determined in the case of unbiased estimates (in x, y coordinates) from this equation

$$\frac{(x-\alpha_1)^2}{\sigma^2_{\alpha_1}} - \frac{2\rho_{\alpha_1,\alpha_2}}{\sigma_{\alpha_1}\sigma_{\alpha_2}}(x-\alpha_1)(y-\alpha_2) + \frac{(y-\alpha_2)^2}{\sigma^2_{\alpha_2}} = 4\left(1-\rho^2_{\alpha_1,\alpha_2}\right). \tag{16.22}$$

If estimates α_1^* and α_2^* are uncorrelated ($\rho_{\alpha_1,\alpha_2} = 0$), then equation (16.22) takes the form

$$\frac{(x-\alpha_1)^2}{(2\sigma_{\alpha_1})^2} + \frac{(y-\alpha_2)^2}{(2\sigma_{\alpha_2})^2} = 1, \tag{16.22a}$$

i. e., ellipese semiaxes equal $2\sigma_{\alpha_1}$ and $2\sigma_{\alpha_2}$.

The dispersion ellipse characterizes the spread of estimates α_1^* and α_2^* relative to the true values of parameters α_1 and α_2.

If the dispersion ellipse corresponding to a given method of estimating parameters α_1 and α_2 is located entirely within the dispersion ellipse corresponding to some other method of estimating the same parameters, then they say that a lesser

spread corresponds to the given estimate method or that the distribution of the estimates of α_1^* and α_2^* found this way has a lesser spread.

If the joint distribution of estimates α_1^* and α_2^* has a lesser spread than the distribution of any other pair of estimates of the same parameters (α_1 and α_2), then estimates α_1^* and α_2^* are called jointly efficient.

Consequently, that method of estimating parameters α_1 and α_2 that insures jointly-efficient estimates result thereby guarantees that minimum-possible spread results.

The following relationships are valid for independent sample values /287 $y_1, \ldots, y_n$ for jointly-efficient estimates:

$$\left.\begin{aligned} \rho_{\alpha_1 . \alpha_2} &= -\frac{M_{\alpha_1 . \alpha_2}\left(\frac{\partial \ln f}{\partial \alpha_1}\frac{\partial \ln f}{\partial \alpha_2}\right)}{\sqrt{\left[M_{\alpha_1}\left(\frac{\partial \ln f}{\partial \alpha_1}\right)^2\right]\left[M_{\alpha_2}\left(\frac{\partial \ln f}{\partial \alpha_2}\right)^2\right]}}; \\ \sigma^2_{\alpha_1} &= \frac{1}{\left(1-\rho^2_{\alpha_1 . \alpha_2}\right) n M_{\alpha_1}\left(\frac{\partial \ln f}{\partial \alpha_1}\right)^2}; \\ \sigma^2_{\alpha_2} &= \frac{1}{\left(1-\rho^2_{\alpha_1 . \alpha_2}\right) n M_{\alpha_2}\left(\frac{\partial \ln f}{\partial \alpha_2}\right)^2}, \end{aligned}\right\} \qquad (16.23)$$

where $f = f(y_i|\alpha_1, \alpha_2)$— unidimensional distribution of sample values y_i for given α_1 and α_2.

Analogous relationships may be obtained also for the case when a statistical link exists among the sample values [23, pgs 539-540].

It follows from formulas (16.23) that, when $\rho_{\alpha_1 . \alpha_2} = 0$, i. e., when

$$M_{\alpha_1 . \alpha_2}\left(\frac{\partial \ln f}{\partial \alpha_1}\frac{\partial \ln f}{\partial \alpha_2}\right) = 0 \qquad (16.24)$$

and given jointly-efficient estimates, the quality of the estimate of each of these parameters, α_1 for example, during simultaneous estimation of two parameters

is the same as if the second parameter α_2 was known and, consequently, would not be subject to estimation.

In other words, when the aforementioned conditions are met, whether or not the second parameter α_2 is known or unknown and subject to simultaneous estimation along with parameter α_1 has no impact on the estimation of given parameter α_1.

However, for existence of jointly-efficient estimates, two conditions analogous to the aforementioned conditions A and B (but more complex) for existence of efficient estimates [23] must be met. Therefore, it may turn out to be impossible in principle to obtain jointly-efficient estimates in a number of cases of practical interest to us. In addition, condition (16.24) may not be met, even for existence of such estimates. Here, as follows from formulas (16.23), variance $\sigma^2_{\alpha_1}$ (or $\sigma^2_{\alpha_2}$) in the case of a simultaneous estimate of parameters α_1 and α_2 is greater than in a case when parameter α_1 is the only unknown.

In many cases, when distribution $L(y_1, \ldots, y_n | \alpha_1, \alpha_2)$ has two unknowns, the task is to estimate the magnitude only of one, α_1 for instance, while α_2 carries no usable information, i. e., is parasitic.

Therefore, it is very important to respond to the next question: is it impossible, given unknown parameters α_1 and α_2, to obtain lesser variance $\sigma^2_{\alpha_1}$ in the estimate of parameter α_1 if there is no requirement here also to /288 estimate parameter α_2, i. e., if one assumes that variance $\sigma^2_{\alpha_2}$ may be as great as desired here?

In the general case, it is possible to obtain such a gain through refusal to estimate the second parameter. However, it was demonstrated in [23] that, in this case, when (given simultaneous estimation of two parameters) resultant estimates (α_1^* and α_2^*) are jointly efficient, refusal to estimate parameter α_2 does not make it possible to decrease+ variance $\sigma^2_{\alpha_1}$ in the parameter α_1 estimate.

+See the page 423 footnote.

This denotes, roughly speaking, that if resultant estimate quality is the best of the possible estimates during simultaneous estimation of parameters α_1 and α_2, then it is impossible to improve upon the best quality of the parameter α_1 estimate by refusing to estimate parameter α_2.

The concepts of asymptotic (where $n \to \infty$)) efficiency and an asymptotically-efficient estimate exist when estimating two parameters analogously to the case of estimating one parameter. In addition, the results achieved above remain (qualitatively) valid also in those cases when a statistical link exists among sample values $(y_1, \ldots, y_n)$.

The examination above for two unknown parameters may be generalized for a case of a random (finite) number of unknown parameters $(\alpha_1, \alpha_2, \ldots, \alpha_m)$.

16.3 Maximum Likelihood Method

The material presented in § 16.2 applies to existence of efficient estimates and makes it possible to compute the magnitude of the efficient estimate variance. However, it does not respond to the question of how it is possible to find efficient or, in any event, sufficiently-good estimates, i. e., how one goes about selecting the form of function $a^*(y_1, \ldots, y_n)$.

There are several methods in mathematical statistics for giving estimates. The most-widespread, especially as applied to problems of receiving signals on a noise background, is the so-called maximum likelihood method. It is the most widespread because other methods either provide in the general case less efficient estimates (for example, the moment method) or is more complicated for realization [for example, the method based on determination of the X-coordinate of the "center of gravity" of distribution $P_y(\alpha)$].

The maximum likelihood method for a case of one unknown parameter α involves the following.

That value of parameter α at which likelihood function $L(y_1, \ldots, y_n|\alpha)$ has its maximum is estimate α^*; consequently, needed estimate α^* is solution of the equation

$$\frac{\partial L(y_1, \dots, y_n \mid \alpha)}{\partial \alpha} = 0, \qquad (16.25)$$

called the likelihood equation. This equation in the general case may have several roots, including roots not depending on (y_1, . . ., y_n). Evidently, roots /289
not depending on (y_1, . . ., y_n) must, a fortiori, be discarded, since it is mandatory that needed estimate α^* be a function of the sample values.

An estimate found by solving equation (16.25) is called a maximum likelihood estimate and designated α_M^*.

In the general case, estimate α_M^* is not the best of the possible estimates and, consequently, may have some bias and no minimum variance. In spite of this, the maximum likelihood method has several very valuable properties. They are especially valuable in those cases when sample values y_1, . . ., y_n are statistically independent. Here, the following postulations demonstrated in mathematical statistics [23] are valid:

1. If efficient estimate α^* exists for parameter α, then maximum likelihood equation (16.25) has a single solution for α^*. In other words, in this case, equation (16.25) has only one root, which is the efficient parameter estimate.

2. If sufficient estimate α^* exists for parameter α, then each likelihood equation root is a function of α^*.

3. For some sufficiently-general conditions [23, pg 56[illegible], conditions 1—3], estimate α_M^* is consistent, asymptotically efficient, and asymptotically normal.

This last property is especially valuable since it occurs for significantly-more general cases than do properties 1 and 2. It follows from this that, given independent samples and $n \gg 1$, the most-likely estimate is the very best—it has essentially zero bias and minimum variance. In addition, since here the estimate law of distribution is very close to normal, variance magnitude completely characterizes this law. This guarantees that the estimate found is very good, not only in the sense of mean-square error magnitude, but also in any other sense.

Finally, a very useful special feature of the maximum likelihood method is that, in this case, finding estimate α^* requires determination only of the position of the function $L(y_1, \ldots, y_n|\alpha)$ maximum with respect to α; therefore, any transformations of the function $L(y_1, \ldots, y_n|\alpha)$ form which do not cause bias (with respect to the α axis) of its maximum do not impact on estimate α_M^*. magnitude and, consequently, on its quality. This circumstance considerably simplifies realization of the device automatically estimating the parameter.

The results presented easily are extended to a case of several unknown parameters $\alpha_1, \ldots, \alpha_m$. Here, estimates $\alpha_{1M}^*, \ldots, \alpha_{mM}^*$ of these parameters are solutions of m likelihood equations of the type

$$\left.\begin{array}{l} \frac{\partial L(y_1, \ldots, y_n|\alpha_1, \ldots, \alpha_m)}{\partial \alpha_1} = 0; \\ \ldots\ldots\ldots\ldots\ldots \\ \frac{\partial L(y_1, \ldots, y_n|\alpha_1, \ldots, \alpha_m)}{\partial \alpha_m} = 0. \end{array}\right\} \tag{16.26}$$

Estimates found in this way for independent sample values $y_1, \ldots, y_n$ /290 have the same useful properties; for example, they are consistent, asymptotically normal, and asymptotically jointly efficient.

16.4 Use of the Maximum Likelihood Method for Reception of Analog Messages on a Noise Background

Let oscillation y(t) at receiver input have the form

$$y(t) = u(\alpha_1, \ldots, \alpha_m; t) + u_{ш}(t), \tag{16.27}$$

where $u_{ш}(t)$ — fluctuating noise, in the general case having a nonstationary process, $\alpha_1, \ldots, \alpha_m$ — unknown signal parameters, which in the general case may be time functions.

Initially, for simplicity we will assume that all these parameters are useful and subject to estimation.

It follows from the preceding section that it is necessary (as a rule) to have a sufficiently-large number n of independent sample values $y_1, \ldots, y_n$

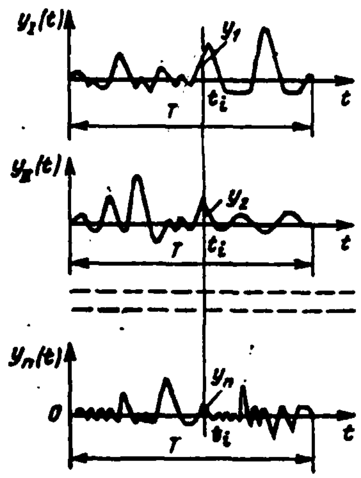


Figure 16.1

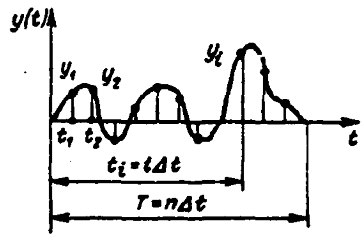


Figure 16.2

of observed magnitude y to use the maximum likelihood method to obtain good-enough estimates. However, in this case, we are concerned with observation of a random process, which also happens to be nonstationary, rather than a random magnitude. Therefore, in the general case, instead of n independent values $y_1, \ldots, y_n$, we must have n realizations $y_I(t), y_{II}(t), \ldots, y_n(t)$ of process y(t) obtained in n independent tests conducted under identical conditions (Figure 16.1).

Here, for each moment $t = t_i$ (Figure 16.1), we will have likelihood functions

$$L(y_1, \ldots, y_n \mid \alpha_1, \ldots, \alpha_m; t),$$

called a simultaneous likelihood function since all sample values $y_1, \ldots, y_n$ included in it are taken from n realizations for the <u>identical moment in time</u> (Figure 16.1).

Solving m likelihood equatio s of the type

$$\left.\begin{array}{l} \dfrac{\partial L(y_1, \ldots, y_n \mid \alpha_1, \ldots, \alpha_m; t)}{\partial \alpha_1} = 0; \\ \cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial L(y_1, \ldots, y_n \mid \alpha_1, \ldots, \alpha_m; t)}{\partial \alpha_m} = 0, \end{array}\right\} \quad (16.28)$$

we will obtain most-likely estimates $\alpha_1^*, \ldots, \alpha_m^*$ for a given moment in time:

$$\alpha_1^* = \alpha_1^*(t); \dots ; \alpha_m^* = \alpha_m^*(t).$$

For the next moment in time, we will obtain a new value of the simultaneous likelihood function, new set of likelihood equations, and new set of estimates $\alpha_1^*, \dots, \alpha_m^*$.

In the general case, when needed parameters $\alpha_1, \dots, \alpha_m$ are time functions /291 and noise is nonstationary, resultant estimates $\alpha_1^*, \dots, \alpha_m^*$ also are time functions and, in addition, the law of distribution of these estimates also will depend on time.

As Grenander demonstrated [115], estimates $\alpha_1^*, \dots, \alpha_m^*$ obtained by solving likelihood equations (16.28) have [given independent $y_1(t), \dots, y_n(t)$ realizations] the useful asymptotic (where $n \to \infty$) properties described in § 16.3, i. e., are consistent, asymptotically normal, and asymptotically jointly efficient.

However, in practical problems of receiving signals on a noise background, the observer (receiver) has at the point of reception usually just one realization of the y(t) process, rather than n independent realizations of this random process. Therefore, of primary interest are cases when good parameter estimates may be obtained from analysis of just one realization y(t) during some finite time T.

Grenander [115] demonstrated that most-likely estimates obtained from analysis of just one realization in interval (0, T) have the same useful asymptotic properties (consistency, efficiency, and normal law of distribution) when $T \to \infty$ if process y(t) is ergodic and have certain general Markov properties.+

As Slepian demonstrated [119], in particular, a case when noise $u_m(t)$ is an ergodic process, while unknown signal parameters $\alpha_1, \dots, \alpha_m$ are constant during the time of observation, boils down to such processes.

+A random process has Markov properties if the last course of this process only to a limited degree depends on its preceding flow. This restriction placed on the nature of the process is formed mathematically in the following manner [11, 115, and others]. It is considered that a process has all the more general Markov properties, the less rigid the restriction is relative to the degree of influence of the process's "history" on its subsequent development.

Since this case is of great practical interest, we will examine it in /292 somewhat more detail. Here, for simplicity, we will limit ourselves to a case of one unknown parameter, i. e., we will assume that

$$y(t)=u(\alpha;\ t)+u_{\text{ш}}(t), \tag{16.29}$$

where $u_{\text{ш}}(t)$ -- noise, being an ergodic process, $u(\alpha; t)$ -- precisely-known signal, with the exception of parameter α subject to estimation and remaining unchanged throughout the entire time of observation (0, T).

There are n sample values $(y_1, \ldots, y_i, \ldots, y_n)$ of total oscillation y(t) during an observation cycle (Figure 16.2).

We will designate:

$$\left.\begin{aligned} u_i &= u(\alpha;\ t_i),\\ u_{\text{ш}i} &= u_{\text{ш}}(t_i);\\ y_i &= y(t_i). \end{aligned}\right\} \tag{16.30}$$

Then, from (16.29), we have

$$\begin{gathered} y_i = u_i + u_{\text{ш}i},\\ i=1,\ 2, \ldots,\ n. \end{gathered} \tag{16.31}$$

Here, the n-dimensional law of distribution of noise $W_{\text{ш}}(u_{\text{ш}1}, \ldots, u_{\text{ш}n})$ is assumed to be precisely kncwn.

Then, in accordance with (16.31), the law of distribution of sample values $(y_1, \ldots, y_n)$ for a given α equals

$$L(y_1, \ldots,\ y_n|\alpha) = W_{\text{ш}}(y_1-u_1, \ldots,\ y_n-u_n),$$

i. e.,

$$\begin{gathered} L(y_1, \ldots,\ y_n\,|\,\alpha) = W_{\text{ш}}[y_1-u(\alpha;\ t_1),\\ y_2-u(\alpha;\ t_2), \ldots,\ y_n-u(\alpha;\ t_n)]. \end{gathered} \tag{16.32}$$

Since $u(\alpha; t_i)$ — known functions of α, then the right side of equality (16.32) is a precisely-known function of sample values (y_1, . . ., y_n) and parameter α; consequently, we know distribution $L(y_1, \ldots, y_n | \alpha)$ and we must find parameter α of this distribution.

This problem coincides completely with the classic problem examined in preceding sections solved in the theory of statistical estimations and, consequently, all conclusions presented in preceding sections are valid for it.

Thus, for example, when $T \to \infty$, it is possible to obtain an infinite number of statistically-independent sample values y_1, . . ., y_n and, consequently, when $T \to \infty$, estimates obtained using the maximum likelihood method must be consistent, asymptotically efficient, and asymptotically normal.

Evidently, all computations performed above for a case of one unknown parameter are valid also for any (finite) number of unknown parameters $\alpha_1, \ldots, \alpha_m$. Consequently, given ergodic noise and constancy [during interval (0, T)] of unknown signal /293 parameters $\alpha_1, \ldots, \alpha_m$, all computations and conclusions made in preceding sections are applicable completely to estimation of these parameters.

Thus, for example, if a jointly-efficient combination $(\alpha_1^*, \ldots, \alpha_m^*)$ of parameter $(\alpha_1, \ldots, \alpha_m)$ estimates exists, then variance $\sigma^2_{\alpha_i}$ of the estimate of any of these parameters α_i may not be decreased due to refusal to estimate all or several remaining unknown parameters.+

This conclusion is especially important in practice since, in many cases, reproduction of various signal parameters has varied significance, while several of the parameters carry no useful information of any kind, i. e., are parasitic. Thus, for instance, in radar it is very important to know whether it is possible to increase accuracy in measurement of target range (i. e., signal lag τ) if one fails simultaneously to measure target radial velocity (i. e., signal frequency f) and vice versa.

+This and conclusions stemming from it apply to a case when signal characteristics are given, i. e., may not be changed, given an estimate refusal.

The theory presented above makes it possible in several cases to answer this question.

For example, let the range and velocity estimate be made using the maximum likelihood method and observation time T be so great that it is possible to obtain a very large number of independent sample values $(y_1, \ldots, y_n)$. Here, as follows from the aforementioned, resultant estimates τ_M^* and f_M^* may with sufficient precision (rising with an increase in T) be considered jointly efficient. Consequently, given such conditions, refusal to measure frequency will not increase range measurement accuracy and vice versa. In other words, it is immaterial from the range measurement point of view whether frequency f is a parasitic unknown parameter or if it is a known parameter subject to measurement.

If a joint efficient combination of estimates (α_1^*, α_2^*) not only exists when there are two unknown parameters α_1 and α_2, but condition (16.24) also is met, i. e., $\rho_{\alpha_1,\alpha_2} = 0$, then the formula for parameter α measurement error variance $\sigma^2_{\alpha_1}$ during simultaneous parameter α_1 and α_2 measurement is identical to that in a case where parameter α_2 is precisely known. Consequently, when the aforementioned conditions are met, the fact that parameter α_2 is unknown or the requirement to measure it does not impact on parameter α_1 measurement accuracy (and vice versa).

These examples show that use of results obtained from mathematical statistics for signal reception problems makes it possible in several cases relatively simply to obtain answers to very-important practical questions.

We now will explain the link between the maximum likelihood method and the maximum inverse probability density method presented in preceding parts of the book.

In the case of the maximum likelihood method, the parameter α value /294 corresponding to the likelihood function $L(y_1, \ldots, y_n|\alpha)$ maximum or, which is the same thing, conditional probability density $P_a(y)$, considered a function of α [see equality (16.1)], is selected as estimate α^*.

If the maximum inverse probability density method is used, that value of

α corresponding to the inverse probability density $P_y(\alpha)$ maximum is selected as the α^* estimate.

But,

$$P_y(\alpha) = kP(\alpha)\,P_\alpha(y), \tag{16.33}$$

where k -- normalizing constant.

If the needed parameter α a priori distribution $P(\alpha)$ is uniform, then

$$P_y(\alpha) = k'P_\alpha(y), \tag{16.34}$$

where k' -- some new value of the normalizing constant not depending on α.

It follows from (16.34) that function $P_y(\alpha)$ and $P_\alpha(y)$ maxima with respect to α coincide. Therefore, the maximum inverse probability density method and the maximum likelihood method completely coincide when the parameter α a priori distribution is uniform.

If distribution $P(\alpha)$ is irregular, then the function $P_y(\alpha)$ and $P_\alpha(y)$ maxima with respect to α do not coincide and, consequently, estimates found using the maximum inverse probability density method and the maximum likelihood method turn out to be different. If a priori distribution $P(\alpha)$ is known, then the estimate found using the maximum inverse probability density method+ turns out to be more efficient since additional a priori information about the signal--a priori distribution $P(\alpha)$ of its parameter, is considered here.

However, as noted in Parts II and III of the book and as will be shown in Chapter 19, given noise in the form of normal white noise (and not only given such noise) and a sufficiently-high signal-to-noise ratio, the type of a priori distribution $P(\alpha)$ essentially does not impact upon the results of the parameter

---

+The maximum inverse probability density method sometimes is called the unconditional maximum likelihood method since, compared to it, the conventional maximum likelihood method, which does not consider a priori distribution $P(\alpha)$, is conditional.

estimate. Therefore, under such conditions, the methods of maximum inverse probability density and maximum likelihood also provide essentially identical results.

If a priori distribution $P(\alpha)$ is unknown, then it is assumed to be uniform in the majority of cases and the function $P_y(\alpha)$ and $P_\alpha(y)$ maxima with respect to α also coincide.

Finally, if it is known that parameter α by nature is not random, then a priori distribution $P(\alpha)$ from the physical point of view loses its meaning and only likelihood function $P_\alpha(y)$ retains meaning. However, since function $P_y(\alpha)$ /295 and $P_\alpha(y)$ maxima with respect to α coincide when $P(\alpha)$ = const, it is possible from the mathematical point of view and when parameter α is not random to use the maximum inverse probability density method if you assume that $P(\alpha)$ = const here.

It follows from what has been said that, in a majority of the interesting cases, methods of maximum inverse probability density and maximum likelihood essentially coincide.

Evidently, everything stated above will apply not only to a case of estimating one parameter, but to cases of estimating several unknown parameters as well.

Examples of the use of the maximum inverse probability density method (and, consequently, the maximum likelihood method as well) for estimation of one parameter were examined in detail in Parts II and III of the book (Chapters 4, 6, 7, and 13).

Use of the maximum likelihood method for simultaneous estimation of two signal parameters was studied in detail for the first time by S. Ye. Fal'kovich [5]. Some results from this study will be presented for illustration in the next section.

16.5 Simultaneous Signal Frequency and Lag Estimation

Let this oscillation be at receiver input

$$y(t)=u_c(t)+u_{\text{ш}}(t),$$

where $u_{\text{ш}}(t)$ -- normal white noise with spectral density N_0; $u_c(t)$ -- signal, which has the form

$$u_c(t)=a(t-\tau)\cos[2\pi(f_0-F)t+\varphi(t)+\varphi_0]; \qquad (16.35)$$

here, φ_0 -- random initial phase with a uniform distribution in the 0 to 2π range, while lag τ and frequency shift F -- unknown parameters subject to estimation. Frequency f_0 and functions $\varphi(t)$ and $a(t - \tau)$ are precisely known (the latter, with the exception of τ). The assumption is that, during time of observation (0, T) τ φ_0, f_0, and F remain unchanged.

It is convenient to write oscillation (16.35) in the following form:

$$u_c(t)=\operatorname{Re}U(t,\ \tau,\ F)e^{j(2\pi f_0 t+\varphi_0)}, \qquad (16.36)$$

where

$$U(t,\ \tau,\ F)=a(t-\tau)e^{j[-2\pi Ft+\varphi(t)]} \qquad (16.37)$$

is complex signal envelope.

In the particular case where F = 0 and $\varphi(t) \equiv 0$, envelope $U(t, \tau, F)$ is a real function.

As will be clear from what follows, accuracy in estimation of τ and F is determined by the type of function

$$\Psi(\tau_1,\ F_1,\ \tau_2,\ F_2)=\frac{1}{2Q}\left|\int\limits_{-\infty}^{\infty}U(t,\ \tau_1,\ F_1)U^*(t,\ \tau_2,\ F_2)\,dt\right|. \qquad (16.38)$$

where U* denotes a complex magnitude conjugate with U, while vertical lines /296 show that the modulus is taken from the corresponding complex magnitude; Q -- signal energy, i. e.,

$$Q = \frac{1}{2} \int_{-\infty}^{\infty} a^2(t)\, dt, \qquad (16.39)$$

Function Ψ attains the greatest value (unity) when $\tau_2 = \tau_1$, $F_2 = F_1$ and decreases (it does not have to be monotonically) by virtue of an increase in the deviation of parameters τ_2 and F_2 from τ_1 and F_1, respectively.

Consequently, Ψ characterizes the link between complex envelopes corresponding to the two different unknown parameter combinations: (τ_1, F_1) and (τ_2 and F_2). Therefore, function Ψ sometimes is called the signal correlation function. However, to avoid confusion with the correlation function used in the theory of random processes, we will call function Ψ a signal function or a function of "correlation."

For the majority of actually-used signals, function Ψ will depend not on values τ_1, F_1 and τ_2, F_2 themselves, but only on differences

$$\tau' = \tau_2 - \tau_1 \quad \text{and} \quad F' = F_2 - F_1. \qquad (16.39a)$$

Here, relationship (16.38) takes the following simplified form:

$$\Psi(\tau', F') = \frac{1}{2Q} \left| \int_{-\infty}^{\infty} U(t, \tau', F')\, U^*(t, 0, 0)\, dt \right|. \qquad (16.40)$$

Assuming that signal function Ψ has the same simplified form, while the signal-to-noise ratio is sufficiently high, Fal'kovich used the maximum likelihood method to obtain the following expressions for parameter τ and F estimate error [5]:

$$\left.\begin{aligned} \overline{\delta_\tau^2} &= \frac{1}{2q\beta_\tau^2} \frac{1}{(1-r^2)}; \\ \overline{\delta_F^2} &= \frac{1}{2q\beta_F^2} \frac{1}{(1-r^2)}; \\ \rho_{\tau,F} &= \frac{\beta_{\tau,F}}{2q\beta_\tau^2 \beta_F^2} \frac{1}{(1-r^2)}, \end{aligned}\right| \qquad (16.41)$$

where

$$\left.\begin{aligned}
&\beta_\tau^2=\left|\frac{\partial^2 \Psi(\tau',0)}{\partial \tau'^2}\right|_{\tau'=0};\\
&\beta_{\tau,F}=\frac{\partial^2 \Psi(\tau',F')}{\partial \tau'\,\partial F'}\bigg|_{\substack{\tau'=0\\F'=0}};\\
&\beta_F^2=\left|\frac{\partial^2 \Psi(0,F')}{\partial F'^2}\right|_{F'=0};\\
&r^2=\frac{\beta_{\tau,F}^2}{\beta_\tau^2\beta_F^2};\\
&q=\frac{Q}{N_0}.
\end{aligned}\right| \qquad (16.42)$$

/297

Here, mean square errors $\overline{\delta_\tau^2}$ and $\overline{\delta_F^2}$ and correlation coefficient $\rho_{\tau,F}$ have exactly the same meaning as in previous sections.+

In the case under examination, estimates τ^* and F^* of parameters τ and F are (asymptotically) unbiased, jointly efficient, and with a normal law of distribution.

Since estimates τ^* and F^* are jointly efficient, refusal to estimate frequency F may not increase parameter τ measurement accuracy and vice versa.** In addition, relationships (16.23) must be valid for jointly-efficient estimates.

If, during the parameter τ estimate, frequency shift F is precisely known or, vice versa, during the frequency shift F estimate, lag τ is known, then one should assume that $\beta_{\tau,F}^2 = 0$ in formulas (16.41) and (16.42).

The result here is:

$$\left.\begin{aligned}
&\overline{\delta_\tau^2}=\frac{1}{2q\beta_\tau^2};\\
&\overline{\delta_F^2}=\frac{1}{2q\beta_F^2};\\
&\rho_{\tau,F}=0.
\end{aligned}\right| \qquad (16.43)$$

+Expressions (16.41) and (16.42) are valid for signals with known amplitude. When signal amplitude is random, then these expressions will remain valid if $\overline{\delta_\tau^2}$ and $\overline{\delta_F^2}$ are understood to be conditional, rather than unconditional, mean squares determined for a given amplitude value (given signal-to-noise ratio q).

**See page 423 footnote.

It follows from comparison of expressions (16.41) and (16.43) that, when $\rho_{\tau,F} = 0$ (or $\overline{\beta^2_{\tau,F}} = 0$), mean square errors $\overline{\delta^2_\tau}$ and $\overline{\delta^2_F}$ during simultaneous measurement of two parameters are determined from the same formulas as is the case when one of these parameters is precisely known.

We will use a signal having the form of a pulse with a bell-shaped envelope as our example

$$a(t-\tau) = a_0 e^{-c(t-\tau)^2} \tag{16.44}$$

and frequency occupation equalling $f_0 + F$, i. e., $\varphi(t) \equiv 0$.

Here, from (16.37) we obtain /298

$$U(t, \tau, F) = a_0 e^{-c(t-\tau)^2} e^{j2\pi Ft},$$

and signal function $\Psi(\tau', F')$, in accordance with (16.40), has the form

$$\Psi(\tau', F') = \frac{1}{2Q}\left|\int_{-\infty}^{\infty} a_0^2 e^{-c(t-\tau')^2} e^{-j2\pi F't} e^{-ct^2} dt\right|.$$

Following simple transformations, we obtain

$$\Psi(\tau', F') = \frac{a_0^2}{2Q} e^{-\frac{c}{2}(\tau')^2} |z|, \tag{16.45}$$

where

$$z = \int_{-\infty}^{\infty} e^{-2c(t-\tau'/2)^2} e^{-j2\pi F't} dt. \tag{16.46}$$

Expression (16.46) may be considered a Fourier transform from integrand

$$f(t) = e^{-2c\left(t-\frac{\tau'}{2}\right)^2}$$

and may be found using conventional methods of operational calculus or from tables. Then, we will obtain

$$z = \sqrt{\frac{\pi}{2c}} e^{-\frac{\pi^2}{2c}(F')^2} e^{-j\pi F'\tau'}$$

and, in accordance with (16.45)

$$\Psi(\tau', F') = \frac{a_0^2}{2Q} e^{-\frac{c}{2}(\tau')^2} \sqrt{\frac{\pi}{2c}}\, e^{-\frac{\pi^2}{2c}(F')^2}. \tag{16.47}$$

But, it follows from (16.39) and (16.44) that

$$Q = \frac{1}{2} \int_{-\infty}^{\infty} a_0^2 e^{-2ct^2} dt = \frac{a_0^2}{2} \sqrt{\frac{\pi}{2c}}. \tag{16.48}$$

Substituting this expression into (16.47), we have

$$\Psi(\tau', F') = e^{-\frac{c}{2}(\tau')^2} e^{-\frac{\pi^2}{2c}(F')^2}. \tag{16.49}$$

As could be expected, the result is

$$\Psi(0, 0) = 1,$$

and, with a τ' and F' increase, magnitude Ψ decreases.

Considering (16.49), formulas (16.42) provide

$$\beta_\tau^2 = c;\ \beta_F^2 = \frac{\pi^2}{c};\ \beta_{\tau,F}^2 = 0.$$

Substituting these results into formulas (16.41), we have /299

$$\left.\begin{aligned} \overline{\delta_\tau^2} &= \frac{1}{2qc}; \\ \overline{\delta_F^2} &= \frac{c}{2q\pi^2}; \\ \rho_{\tau,F} &= 0. \end{aligned}\right\} \tag{16.50}$$

Since, the result is $\rho_{\tau,F} = 0$ in this case, then mean square errors $\overline{\delta_\tau^2}$ and

$\overline{\delta_F^2}$ are determined in the identical manner as in the case of estimation of one of these parameters when the second parameter is precisely known.+

It follows from formula (16.50) that, in this case, mean-square errors $\overline{\delta_\tau^2}$ and $\overline{\delta_F^2}$ will depend only on signal-to-noise power ratio q and on parameter c, which characterizes pulse duration [see expression (16.44)] and, consequently, the width of its spectrum: the greater the c, the shorter the pulse duration and the broader its spectrum. Therefore, it is quite understandable that, with an increase in c, the mean square lag τ measurement error decreases, while the mean-square frequency shift F measurement error increases.

The reader will find the principles of constructing an optimum receiver for simultaneous measurement of two parameters, as well as additional problems involving optimum estimation of signal parameters, in Fal'kovich's book [5]. We will dwell briefly only on one of the most-important problems--the influence of signal shape during simultaneous measurement of two parameters. Here, for specificity, we will introduce an examination applicable to radar, i. e., we will assume that lag τ is a measure of an object's (aircraft and so on) range, while frequency shift F is a measure of this object's radial velocity.

This problem was examined in several works, including those of Woodward [2], Siebert [103], and in the aforementioned Fal'kovich book [5]. Several fundamental results of this examination will be presented in the next section.

+However, it still does not follow from this that, when $\rho_{\tau,F} = 0$, the necessity for simultaneous parameter F measurement may not in any way influence the accuracy achieved in measurement of parameter τ (and vice versa).

Actually, it follows from formula (16.50) that, for given signal-to-noise ratio q, a decrease in error $\sqrt{\overline{\delta_F^2}}$ requires a decrease in the magnitude of coefficient c (i. e., an increase in pulse duration). But, here, the error $\sqrt{\overline{\delta_\tau^2}}$ magnitude will rise. Therefore, if the coefficient c increase during practical system accomplishment is restricted by this same error $\sqrt{\overline{\delta_\tau^2}}$ rise above the permissible level (and not by any other circumstances, such as design approaches), then the necessity for sufficiently-accurate parameter F measurement may lead to an increase in parameter τ measurement error (and vice versa). Consequently, when $\rho_{\tau,F} = 0$, this hidden influence of the necessity for parameter F measurement on parameter τ measurement accuracy (and vice versa) may occur.

16.6 Impact of Signal Shape During Simultaneous Frequency and Lag Measurement /300

As follows from formulas (16.41) and (16.42), parameter τ and F measurement accuracy will depend, given signal-to-noise power ratio q, only on the type of signal function $\Psi(\tau', F')$. Therefore, one of the basic (if not the basic) optimizations of the signal $u_c(t)$ shape is the proper type of signal function $\Psi(\tau', F')$.

Above plane (τ', F'), this function forms a surface, whose vertex is located at the origin of the coordinates and which equals unity since always

$$\Psi(0, 0) = 1. \tag{16.51}$$

By virtue of the τ' and F' increase, function $\Psi(\tau', F')$ decreases monotonically or not monotonically, depending on signal $u_c(t)$ type. In the latter case, surface $\Psi(\tau', F')$ has, besides a central vertex, other additional vertices of lesser magnitude. However, for any signal type, the complete volume formed by surface $\Psi(\tau', F')$ with plane (0, τ', F;) turns out to be identical and equalling unity, i. e., always

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Psi(\tau', F')\, d\tau'\, dF' = 1. \tag{16.52}$$

In radar, this relationship is called the ambiguity principle. [The validity of relationship (16.52) is demonstrated in general form by substitution of expression (16.40) into formula (16.52) and performance of the corresponding mathematical transforms]. The same procedure is used for convenience in geometric interpretation of surface $\Psi(\tau', F')$ and the volume it forms as is used to depict mountain peaks on a map--sections of surface $\Psi(\tau', F')$ with horizontal planes are drawn at different levels ("altitudes") and sections corresponding to different levels are noted on plane (0, τ', F'), using different densities of shading for instance; the higher the level, the denser the shading.

Usually, three gradations of magnitude Ψ are used for an approximate /301 characterization of the type of surface $\Psi(\tau', F')$, namely: solid sections of plane (0, τ', F') correspond to regions of high "correlation," i. e., $\Psi(\tau', F') \approx 1$; shading corresponds to slight "correlation," i. e. to regions where

$0 < \Psi(\tau', F') \ll 1$; finally, light sections of plane $(0, \tau', F')$ apply to the "uncorrelated" region where $\Psi(\tau', F') = 0$.

Formula (16.49) describes the signal function for a signal in the form of a sinusoidal pulse with a bell-shaped envelope and surface section $\Psi(\tau', F')$ of

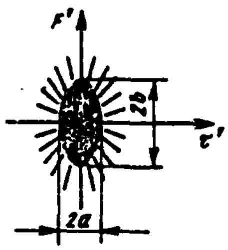


Figure 16.3

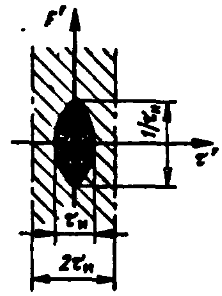


Figure 16.4

a horizontal plane has the form of an ellipse with semi-axis ratio

$$\frac{b}{a} = \frac{c}{\pi}. \tag{16.53}$$

Therefore, Figure 16.3 characterizes the type of function $\Psi(\tau', F')$ for such a signal.

Figure 16.4 corresponds to a sinusoidal pulse with a rectangular envelope with duration $\tau_и$.

A train of four such coherent pulses (Figure 16.5a) corresponds to Figure 16.5b. Evidently, surface $\Psi(\tau', F')$, having one vertex and no region of zero "correlation" [since, in this case, $\Psi(\tau', F')$ equals zero only when $\tau' \to \infty$ or $F' \to \infty$], corresponds to Figure 16.3. A surface with one vertex having a region both with large and small, as well as with zero, "correlation" (the latter occurs when $|\tau'| > \tau_и$) corresponds to Figure 16.4.

Finally, a surface with a large number of vertices and with regions /302 with a large, small, and zero "correlations" corresponds to Figure 16.5. Appropriate figures may be drawn for other signal types as well.

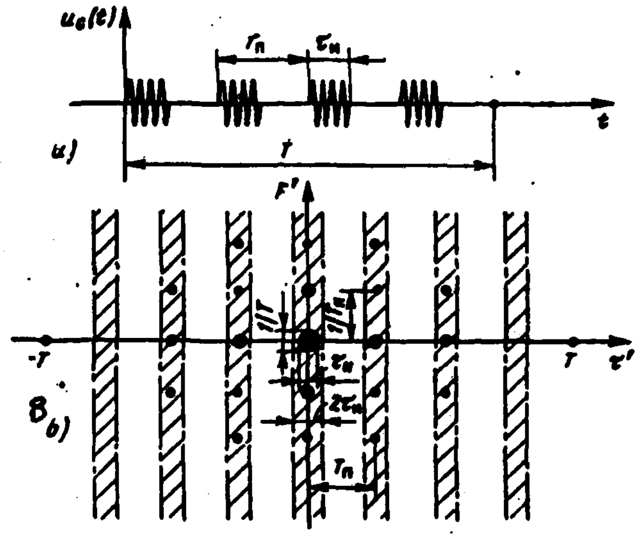


Figure 16.5

When solving the problem of which type of function $\Psi(\tau', F')$ [and, consequently, of signal $u_c(t)$] is the best, one must consider a series of circumstances, with the following being the most important:

1. The requirement for high parameter τ and F measurement accuracy and the absence of measurement ambiguity.

2. The requirement for high resolution.

3. Possible simplification of the measurement system (radar).

It is easy to become convinced that the following conditions must be met in or ɿr to increase parameter τ and F measurement accuracy and eliminate function $\Psi(\tau', F')$ ambiguity:

a) Only one (central) region of high "correlation" must exist within a given range of possible measured magnitude (τ and F) values.

b) This region must be as small as possible.

The first of these conditions is required to eliminate measurement ambiguity,

while the second is needed to increase measurement accuracy. The following computations confirm the validity of these postulations.

Insuring high parameter τ and F measurement accuracy in the presence of noise requires that even very slight changes in these parameters cause significant changes in signal $u_c(t, \tau, F)$ shape since, otherwise, the probability is great that parameter changes will not be detected due to the changing noise action. But, it is evident from relationships (16.39a) and (16.40) that, given absolutely no parameter changes (where $\tau_2 = \tau_1$ and $F_2 = F_1$, i. e., when $\tau' = 0$ and $F' = 0$), the result is

$$\Psi(\tau', F') = 1.$$

Consequently, zero or infinitely-slight parameter changes, which in principle are impossible to detect even given the slightest possible noise, correspond to Ψ values equalling or infinitely-close to unity. Therefore, the region of function $\Psi(\tau', F')$ values in a range where $\Psi(\tau', F') \approx 1$, i. e., a region of high "correlation," is called the ambiguity region.

The smaller the ambiguity region, i. e., the region of high "correlation" turns out to be, the higher the parameter τ and F measurement accuracy.

Hence, it also follows that, if not one, but several regions of high "correlation," i. e., ambiguity regions, fall within the possible measured parameter values, then there is ambiguity in determination of τ and F magnitude. Thus, for example, in the case of a pulse train (Figure 16.5), measurement ambiguity arises if τ' or F' depart bounds $\pm(T_n - \tau_u)$ and $\pm\left(\frac{1}{T_n} - \frac{1}{2T}\right)$, respectively.

We now will explain how function $\Psi(\tau', F')$ type [i. e., signal $u_c(t)$ type] influences measurement system (radar) resolution. /303

First, we will examine the simplest case when there is a total of two objects in space--a given object with parameters (τ_1 and F_1) and interfering object with parameters (τ_2 and F_2). We will assume that signals $u_c(t, \tau_1, F_1)$ and $u_c(t, \tau_2, F_2)$ reflected from both objects differ only in the values of parameters τ and F. Then, it again is possible to form a (16.40)-type "correlation" function for these signals and the figures plotted on its basis, such as Figures 16.3, 16.4, and 16.5, turn out to be valid.

If both objects have identical parameter τ and F values (i. e., $\tau' = 0$ and $F' = 0$), then they in principle are indistinguishable. Here, as follows from (16.40), the result is $\Psi(\tau', F') = 1$.

Consequently, if function $\Psi(\tau', F')$ equals or is infinitely-close to unity, then both objects are indistinguishable, regardless how slight the noise is.

Hence, it follows that regions of high "correlation" [where $\Psi(\tau', F') \approx 1$] are ambiguity regions, not only from the point of view of the impossibility of accurate measurement of the parameters of a single object, but from the point of view of the impossibility (in this region of parameters) of distinguishing one object from another, i. e., from the point of view of system resolution. Therefore, it also is desirable from the resolution standpoint that the region of "correlation" be as narrow as possible.

The lower the function $\Psi(\tau', F')$ value for signals from two objects, the easier it is to distinguish them from each other in the presence of noise and, consequently, the higher the system resolution.

As pointed out above, $\Psi(\tau', F') = 0$ is the case for orthogonal signals. Therefore, the solution turns out to be complete for such signals--it is possible completely to avoid the interference from signal $u_c(t, \tau_2, F_2)$ when signal $u_c(t, \tau_1, F_1)$ is being received (if orthogonality is achieved by the signals not overlapping over time or with respect to frequency spectrum, then such a complete solution is achieved through time or frequency gating, respectively).

When, instead of one, there are several interfering signals, the problem of signal shape [or of function $\Psi(\tau', F')$ type] influence on system resolution

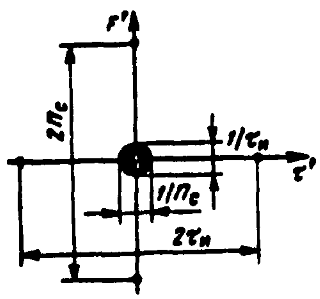


Figure 16.6

is complicated greatly. In particular, function $\Psi(\tau', F')$ type in regions of slight "correlation" (in the shaded regions of Figures 16.3--16.5) acquires great significance. However, in this case also, the ideal signal would be the one (Figure 16.6) in which the region of high "correlation" would be as small as possible, while there would be no region of slight "correlation" (in Figure 16.6, $\tau_{и}$ denotes signal duration, while Π_c -- width of its spectrum; $\Pi_c \tau_{и} \gg 1$). Evidently, such a signal is ideal also from the standpoint of greatest measurement accuracy and absence of ambiguity. However, obtaining the ideal signal encounters serious principle and practical difficulties.

The principle difference involves relationship (16.52) presented above, /304 i. e., the "ambiguity principle," from which it follows that the complete volume beneath surface $\Psi(\tau', F')$ always equals unity.

Actually, obtaining the ideal signal (Figure 16.6) means making that part of the region beneath surface $\Psi(\tau', F')$, located near the origin of the coordinates as small as possible. But, since the complete volume beneath surface $\Psi(\tau', F')$ must remain unchanged (equal to unity), then an increase in part of the volume located outside the central region of high "correlation" unavoidably increases here. This signifies that, when there is a decrease in the central region of high "correlation," the level of "correlation" outside of this region must be increased and additional regions of high "correlation" even may appear. This may lead to a deterioration in resolution or onset of ambiguity (if the additional regions of high "correlation" turn out to be within the range of possible measured parameter values). Practical problems of obtaining a signal that is close to ideal involve equipment complexity.

It is pointed out in [103] that noise-like signals or signals in the form of a coded pulse train fall in the category of signals with a function $\Psi(\tau', F')$, close to ideal. Here, a coded signal is obtained, for example, from an initial pulse of duration $\tau_{и}$ with sinusoidal occupation, which will be divided into $\Pi_c \tau_{и}$ intervals, each with duration $\frac{1}{\Pi_c}$ (here, $\tau_{и}$ -- duration of the entire signal, while Π_c -- width of its spectrum). Then, rf oscillations during transition from one given interval to the next are retained unchanged or are rotated in phase 180°--in accordance with whether the corresponding code character in binary code comprising $\Pi_c \tau_{и}$ code characters equals zero or unity. The selected code is such

that the "correlation" function $\Psi(\tau', F')$ minimum is insured for all τ' and F' values that differ from zero.

In conclusion, the following, very important, circumstance that has not had suffcient attention devoted to it in the literature should be underscored. All conclusions made above concerning the influence of the type of signal function $\Psi(\tau', F')$ (and, consequently, the signal type as well) on parameter τ and F measurement accuracy completely are valid only given sufficiently-high signal-to-noise ratio q.

This is because, first, only in this case are initial formulas (16.41) and (16.42) valid. Therefore, only in the event of a high signal-to-noise ratio is it possible to affirm that, for a given signal-to-noise power ratio, parameter τ and F measurement accuracy completely is determined by the type of signal function $\Psi(\tau', F')$. /305

Second, it is possible from the following computations to become convinced that the above conclusions may be completely invalid for a low signal-to-noise ratio.

It was noted above that, in the case of noise-like and coded signals, it is possible from the measurement accuracy and ambiguity standpoint to come close to the ideal case depicted in Figure 16.6. But, it follows from this figure that, when $\Pi_c\tau_u \to \infty$, the region of high "correlation" will strive towards zero. This signifies that it is possible, using an unlimited increase in the number of signal $\Pi_c\tau_u$ degrees of freedom, to make a parameter τ and F measurement error as slight as desired, even while keeping signal-to-noise power ratio q unchanged and finite. Therefore, if you assume that this result is valid for any magnitude q, then this will denote that it is possible by means of an unlimited increase in $\Pi_c\tau_u$ to bring the measurement error to zero, even for a very low (but finite) signal-to-noise ratio.

But, this result contradicts fact since it is impossible, given low signal-to-noise ratio q, regardless of the signal $u_c(t)$ shape, to solve even the signal/no signal at receiver input problem reliably enough and, consequently, it is impossible even to guarantee that τ and F measurement results will apply to a usable signal and not to noise, i. e., regardless of the degree to which they are usable.

Actually, solution of the signal/no signal at input problem is nothing but binary signal detection. It was demonstrated in Chapters 5 and 9 that, for simple binary signal detection, detection error probabilities will depend exclusively on the signal-to-noise power ratio [see, formulas (5.29), (9.49), and (9.83, for example] and, when the value of this ratio is low, reliable detection is impossible, regardless of signal shape. In the example of parameter τ and F measurement examined, we are dealing with complex, not simple, binary detection (since parameters τ and/or F are unknown during signal detection). But, it is evident that, if the given signal-to-noise power ratio is insufficient for reliable simple detection, then it is all the more insufficient for the same level of complex detection reliability. Therefore, it is evident that, when q is low, achievement of high detection reliability and/or high parameter τ and F measurement accuracy in principle is impossible, regardless of signal shape.

It follows from what has been stated that signal shape complexity (increasing its number of degrees of freedom $\Pi_c \tau_{\text{и}}$) for the purpose of decreasing the region of high "correlation" (Figure 16.6) and increasing measurement accuracy actually may provide an increase in accuracy only when signal-to-noise ratio q is sufficiently high. Here, the greater the number of degrees of freedom $\Pi_c \tau_{\text{и}}$, the greater magnitude q must be in order for the results obtained to be valid.

Here, we have an analogy with a simpler case--measurement of one parameter (x), examined in Chapter 6. Here, it was demonstrated from the analytical results Kotel'nikov obtained and geometric interpretation that measurement accuracy /306 increases without restraint when the number of signal degrees of freedom increases, but signal-to-noise ratio q required for realization of this accuracy rises simultaneously--otherwise, anomalous errors unavoidably arise [a skip from one multidimensional spiral winding (Figure 6.8) to another] and, consequently, accuracy deteriorates sharply.

The problem of the impact of signal shape on radio reception noise immunity is examined in some detail in many works [5, 127—129, 135, 189—191, and others]. However, the examination considers anomalous errors only in several of them [190, 191, and others].

CHAPTER SEVENTEEN

/306

GENERALIZED OPTIMIZATIONS (USE OF THE THEORY OF STATISTICAL DECISIONS)

17.1 Problem Formulation

The examination in preceding chapters applied to more-widespread, but particular, receiver optimizatons--to criteria of minimum composite error probability, minimum root-mean-square error, and others. However, in several cases, selection of a specific optimization type turns out to be difficult. Therefore, it is very important to know how the selected optimization type impacts on optimum receiver structure and properties. This means that it is necessary to introduce generalized optimizations, which would encompass complete classes of particular criteria. This approach turns out to be possible when the theory of statistical decisions is used.

The theory of statistical decisions, the foundations for which were laid by Wald [30, 16, 118], is a development and generalization of methods of testing statistical hypotheses and estimating distribution parameters and, to a significant degree, uses results from game theory. Therefore, such terminology characteristic of game theory as "decision," "loss" (or "deficit"), "risk," and so on are used, in particular, in the theory of statistical decisions.

In accordance with this theory, the problem of optimum reception of a signal on a noise background is formulated in the following way.

Mixture (not mandatorily additive) y(t) of signal $u_c(t)$ and noise $u_ш(t)$, arrives at receiver input (Figure 17.1), i. e.,

$$y(t) = u_c(t) \otimes u_ш(t).$$

Here, the character $\otimes$ denotes "mixture."

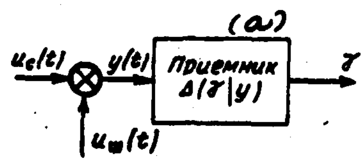


Figure 17.1. (a) -- Receiver.

In a particular case, given additive signal and noise, this mixture is /307 a sum, i. e.,

$$y(t) = u_c(t) + u_ш(t).$$

Noise statistical characteristics are assumed to be precisely known, i. e., the n-dimensional law of noise distribution $W_ш(u_{ш1}.$ is known. The signal may be represented in the following form:

$$u_c(t) = u_c(t;\ x;\ \alpha_1, \ldots, \alpha_m), \tag{17.1}$$

where x, $\alpha_1, \ldots, \alpha_m$ -- signal parameters unknown at the point of reception. These parameters during time of observation (0, T) may be unknown constant magnitude or unknown time functions x(t); $\alpha_1(t), \ldots, \alpha_m(t)$.

Here, x is a usable message subject to reproduction, while $\alpha_1, \ldots, \alpha_m$ -- parasitic parameters, i. e., parameters containing no information concerning the reproduced message.

Unknown magnitudes (or time functions) x_1, $\alpha_1, \ldots, \alpha_m$ at the point of reception

are considered random magnitudes (or random time functions) having known a priori distributions P(x) and $P(\alpha_1, \ldots, \alpha_m)$.

Since parasitic parameters $\alpha_1, \ldots, \alpha_m$ do not comprise any information about message x, there is no statistical link between these parameters and message x. It is assumed that signal $u_c(t)$ dependence on time t and parameters $(x; \alpha_1, \ldots, \alpha_m)$ at the point of reception are precisely known; consequently, given known parameters $(x; \alpha_1, \ldots, \alpha_m)$, signal $u_c(t)$ is precisely known. This signal is called a signal with unknown or random parameters. The theory presented below also is valid for completely-random signals (a priori n-dimensional distributions known at the point of reception) [120]. However, since signals with unknown parameters are more typical in radio engineering, further exposition applies to a (17.1)-type signal.

Since signal and noise statistical characteristics are assumed to be precisely known, the following joint distribution is known or may be ac rately computed

$$P(x, y) = P(x) P_x(y). \qquad (17.2)$$

This distribution more simply will be found, given additive signal and noise. In this case

$$y(t) = u_c(t; x; \alpha_1, \ldots, \alpha_m) + u_m(t); \qquad (17.3)$$

therefore

$$P_{x, \alpha_1, \ldots, \alpha_m}(y) = W_m(y - u_c) \qquad (17.4)$$

and /308

$$P_x(y) = \int_{A_{\alpha_1}} \ldots \int_{A_{\alpha_m}} P_{x, \alpha_1, \ldots, \alpha_m}(y) P(\alpha_1, \ldots, \alpha_m) d\alpha_1 \ldots d\alpha_m. \qquad (17.5)$$

Consequently,

$$P_x(y) = \int_{A_{\alpha_1}} \ldots \int_{A_{\alpha_m}} W_m(y - u_c) P(\alpha_1, \ldots, \alpha_m) d\alpha_1 \ldots d\alpha_m, \qquad (17.6)$$

where

$$u_c = u_c(t;\ x;\ \alpha_1, \dots, \alpha_m).$$

Since distributions $W_m(u_m)$ and $P(\alpha_1, \dots, \alpha_m)$ are assumed to be known, then formula (17.6) makes it possible to determine distribution $P_x(y)$. Following this, joint distribution $P(x, y)$ may be found from formula (17.2).

So, signal-plus-noise $y(t)$, whose statistical characteristics are precisely known, arrives at receiver input (Figure 17.1). The receiving device analyzes oscillation $y(t)$ during time of observation $(0, T)$ and, based upon this analysis, supplies decision γ at receiver output. The operations performed in the receiver with oscillation $y(t)$ to form decision γ are called a decision rule and are designated $\Delta(\gamma \mid y)$. The type of decision γ will depend on receiver purpose.

If the task is just binary signal detection, then γ has a total of two discrete values γ_0 ("no signal") and γ_1 ("signal").

For simple reproduction of discrete messages $x_1, \dots, x_m$, decision γ has corresponding discrete values $\gamma_1, \dots, {}_m$. For simple reproduction of analog message x or $x(t)$, decision γ is analog magnitude γ or time function $\gamma(t)$, respectively. Here, in a case of precise reproduction

$$\gamma = x, \qquad \text{or} \qquad \gamma(t) = x(t).$$

In several cases, the requirement is not simple message reproduction, but rather reproduction with performance of some additional operations such as amplification, differentiation, integration, prediction of the future law of message change, and so on. Thus, for example, the requirement in radar often is to use observation of target trajectory over given time interval $(0, T)$ as the basis for a prediction of a type of trajectory outside the bounds of this interval (for proper antiaircraft shell direction).

The following condition must be met in these cases when there is *no* noise

$$\gamma(t) = Dx(t),$$

where D — some operator corresponding to the requisite transform (differentiation, prediction, and so on).

Decision rule $\Delta(\gamma|y)$ may be <u>regular (non-randomized)</u> or <u>statistical (randomized)</u>.

A regular* rule is one in which precisely-defined (i. e., with probability /309 equalling unity) decision γ [or γ(t)] corresponds to each signal-plus-noise realization y(t).

A statistical (irregular) rule is one in which decision γ is linked with y by a statistical, rather than regular, dependency, i. e., for a given y, the known is *not* decision γ, which the receiver will make, but only the probability of that decision.

All contemporary receiving devices operate in accordance with a regular law (only equipment instability and malfunctions cause deviations from regularity). However, if a special set of statistical mechanisms are installed in a receiver, it is possible to force it to operate with respect to a statistical law.

These same statistical decision rules are inherent in all living organisms, including a human. The reaction of a given person to a given specific action of the external environment may not be predicted with complete validity--only the probabilities of certain reactions to a given action may be known beforehand. Presence of such irregularity in the conduct of living organisms turns out to be harmful for these organisms in some cases and a useful factor in others. It is useful, in particular, in that it hampers its enemies struggling with this organism.

Correspondingly, in machinery as well, particularly in receiving devices,

---

*Here and in future, the terms regular and irregular are synonyms for the terms determinate and indeterminate, respectively.

irregular operation in many cases will hinder results, but may turn out to be useful under certain complex conditions. Thus, for example, irregular operation of radar receivers designed for employment against enemy aircraft may hinder the enemy in creating effectively-organized electronic countermeasures.

Thus, in a majority of cases, it is advisible to design receivers with regular decision rules but, in some special cases, it may turn out to be more useful to have receivers with statistical decision rules.

If a decision rule is statistical, then $\Delta(\gamma|y)$ denotes the probability that the receiver will make decision γ when realization y(t) is present at its input. If γ is an analog magnitude, then $\Delta(\gamma|y)$ is not a probability, but a decision γ probability density. If γ is a time function, then $\Delta(\dot{\gamma}|y)$ is a multidimensional (n-dimensional) probability density.

This regular dependence exists between realization y(t) and decision γ, given a regular decision rule

$$\gamma=\Gamma(y); \tag{17.7}$$

therefore, expression (17.7) expresses a regular decision rule.

If, in the case of a regular decision rule also, $\overline{\Delta(\gamma|y)}$ is understood for the purposes of standardization to be the probability density of decision γ /310 for a given y, then it should be assumed that

$$\Delta(\gamma|y)=\delta[\gamma-\Gamma(y)], \tag{17.8}$$

where $\delta(z)$ is the delta-function of z.

Thus, for all decision rules (regular and irregular), it is possible to understand decision rule $\Delta(\gamma|y)$ to be the probability (or probability density) that decision γ will be made for a given y. Here, in the case of regular decision rules, $\Delta(\gamma|y)$ is a (17.8)-type delta-function.

Thus, based on analysis of realization y(t), in accordance with rule $\Delta(\gamma|y)$,

a receiver processes decision γ. Evidently, decision rule $\Delta(\gamma|y)$ is the optimum rule (i. e., that receiver structure) in which decision γ turns out to be best in some sense. Therefore, the quantitative characteristic of decision γ quality should be selected in order to find the optimum decision rule.

A loss function ("deficit" function) is that characteristic $l(x, \gamma)$, designating losses ("deficits") corresponding to each combination of message x and decision γ taken. A quadratic loss function of the following type is the most widespread during simple reproduction

$$l(x, \gamma)=(\gamma-x)^2. \quad (17.9)$$

In the general case, function $l(x, \gamma)$ may have the most-varied type depending on receiver purpose (detection, simple reproduction, prediction, and so forth) and the requirements levied on it. However, in all cases, it must characterize losses linked with a given message x and decision γ combination, i. e., the less favorable (from the receiver purpose standpoint) given combination (x, γ) is, the greater the magnitude $l(x, \gamma)$ corresponding to it must be.

In a case of simple reproduction, an increase in the reproduction error modulus, $|\gamma-x|$, manifests itself in deterioration of decision quality. Therefore, during simple reproduction, a (17.9)-type quadratic function is one of the most suitable.

In several cases, a loss function may depend not only on current x and γ values, but also on all values of x and γ in a certain region of change of their arguments. In this case, the loss function is converted to a functional.

Finally, for some loss function $l(x, \gamma)$ types, it turns out to be dependent also on decision rule $\Delta(\gamma|y)$.

For example, let the following function be the loss function:

$$l(x, \gamma)=-\log P(x|\gamma)=\log\frac{1}{P(x|\gamma)}, \quad (17.10)$$

where $P(x|\gamma)$ -- probability (or probability density) x for a given γ. The lower probability $P(x|\gamma)$, , the greater the ambiguity x for a given γ. Therefore, in

information theory, expression (17.10) is called ambiguity x for a given γ. /311 Evidently, the greater this ambiguity, the worse things are; therefore, a (17.10)-type function in principle is useful for an estimate of receiver quality and may be selected as the loss function. Since probability x for a given γ will depend on receiver structure, i. e., on decision rule $\Delta(\gamma|y)$, then a (17.10)-type loss function also will depend on decision rule $\Delta(\gamma|y)$. On the other hand, a (17.9)-type function is an example of a loss function that does not depend on decision rule $\Delta(\gamma|y)$.

Thus, loss function $l(x, \gamma)$ suitable for the operating conditions of a given receiver is selected when estimating receiver quality.

Since x and γ are random magnitudes (or random time functions), then loss $l(x, \gamma)$ also is a random magnitude. Therefore, decision quality in the theory of statistical decisions is evaluated not by the magnitude of loss $l(x, \gamma)$, but by its expected value:

$$R=\overline{l(x, \gamma)}, \tag{17.11}$$

where a superimposed line denotes statistical averaging, i. e.,

$$R=\int_{A_x} dx \int_{A_\gamma} l(x, \gamma)\, P(x, \gamma)\, d\gamma, \tag{17.12}$$

where A_x and A_γ -- regions of all possible x and γ values.

Magnitude R, called the <u>average loss</u> (average deficit) is a measure of decision quality (receiver quality)--the smaller this magnitude, the better the decision. Therefore, that decision rule $\Delta(\gamma|y)$, which insures the minimum R magnitude is called the best (optimum) rule. Decision rules $\Delta^*(\gamma|y)$ providing the minimum R magnitude are called Bayes decision rules.

If loss function $l(x, \gamma)$ is such that it will not depend on decision rule $\Delta(\gamma|y)$, then magnitude $l(x, \gamma)$ is called <u>risk</u>, while R is the <u>average risk</u>.

If loss function $l(x, \gamma)$ will depend on decision rule $\Delta(\gamma|y)$, then finding the R minimum, i. e., seeking the optimum decision rule, becomes much more difficult.

Therefore, in a majority of cases, the loss function is selected so that it will not depend on the decison rule. In future, we will limit ourselves for simplicity to just such cases and accordingly we will call magnitude R the average risk. Here, the optimum decision rule is the one providing minimum average risk $\Delta(\gamma|y)$ i. e., converts expression (17.12) into a minimum. It follows from this that such a decision rule falls in the Bayes decision rule class. Seeking this rule, we thereby will find a mathematical description of the optimum receiver structure.

Characterization of message fidelity by <u>one</u> number--average risk R--naturally is not universal or optimum in all respects and generally it is impossible to create a criterion that is standard and optimum in all respects.

However, the minimum average risk criterion is significantly more general /312 than the criteria presented in preceding chapters and they may be obtained from it as particular cases.

17.2 Basic Relationships for the Minimum Average Risk Criterion

Expression (17.12) for average risk is valid for various types of messages x and decisons γ, with the following comments:

1. In a case of discrete x and γ, the integrals in this expression must be replaced by sums, while $P(x, \gamma)$ should be understood to be the <u>composite probability</u> of magnitudes x and γ.

2. In a case of analog x and γ, distribution $P(x, \gamma)$ is the x and γ composite probability <u>density</u>.

If x and γ are random time functions, then distribution $P(x, \gamma)$ is n-dimensional.

We will examine several examples for illustration.

For example, let a message have two possible values $x = x_0$ and $x = x_1$ (binary transmission) and there be only two possible decisions $\gamma = \gamma_0$ (message x_0 is at input) and $\gamma = \gamma_1$ (message x_1 is at input).

In a case of telegraph radio communications, x_0 and x_1 should be understood to mean spacing and pulsing, while decisions γ_0 and γ_1 mean supplying spacing and pulsing, respectively, at output.

In a case of detecting signal presence or absence, values x_0 and x_1 correspond to signal absence and presence, while γ_0 and γ_1 to the decisions "no" (no signal at input) and "yes" (signal at input).

In this case, expression (17.12) takes the form of a sum:

$$R=\sum_{x}\sum_{\gamma} l(x,\gamma)P(x,\gamma). \tag{17.13}$$

In the case under examination, the following are possible message x and decision γ combinations:

$$(x_0,\gamma_0),(x_0,\gamma_1),(x_1,\gamma_0) \quad \text{and} \quad (x_1,\gamma_1). \tag{17.14}$$

Evidently, the first and fourth combinations correspond to correct decisions, while the second and third correspond to incorrect decisions.

We will designate losses corresponding to these combinations:

$$l(x_0,\gamma_0)=c_1;\ l(x_0,\gamma_1)=c_2;\ l(x_1,\gamma_0)=c_3;\ l(x_1,\gamma_1)=c_4. \tag{17.15}$$

Then, in accordance with (17.13), we have

$$R=c_1P(x_0,\gamma_0)+c_2P(x_0,\gamma_1)+c_3P(x_1,\gamma_0)+c_4P(x_1,\gamma_1). \tag{17.16}$$

Usually, zero losses are ascribed to correct decisions, i. e., we assume:

$$c_1=0;\quad c_4=0; \tag{17.17}$$

then /313

$$R=c_2P(x_0,\gamma_1)+c_3P(x_1,\gamma_0). \tag{17.18}$$

This expression may be written also in the following form:

$$R = c_2 P(x_0) P_{x_0}(\gamma_1) + c_3 P(x_1) P_{x_1}(\gamma_0). \qquad (17.19)$$

For instance, let's examine a case of binary signal detection. Then, evidently,

$$P_{x_0}(\gamma_1) = P_{\text{лт}} \quad \text{and} \quad P_{x_1}(\gamma_0) = P_{\text{пр}}, \qquad (17.20)$$

where $P_{\text{лт}}$ and $P_{\text{пр}}$ -- false-alarm and signal miss probabilities.

Here, average risk R equals

$$R = c_2 P(x_0) P_{\text{лт}} + c_3 P(x_1) P_{\text{пр}}. \qquad (17.21)$$

This expression completely coincides with expression (14.12) since c_2 and c_3 play the role of weight factors a and b.

If you assume that

$$c_2 = c_3 = 1. \qquad (17.22)$$

then, the result is

$$R = P(x_0) P_{\text{лт}} + P(x_1) P_{\text{пр}} = P_{\text{ош}}, \qquad (17.23)$$

i. e., for losses c_1, c_2, c_3, and c_4 selected in accordance with relationships (17.17) and (17.22), average risk R coincides with composite error probability $P_{\text{ош}}$ and, consequently, the minimum average risk criterion coincides with the minimum composite error probability criterion.

Consequently, criteria examined in Chapter 14 actually are particular cases of the minimum average risk criterion.

We now will examine an example of simple reproduction of an analog message. Here, as indicated in § 17.1, a (17.9)-type quadratic function is one of the most-widespread loss function $l(x, \gamma)$ types.

Here, from (17.9) and (17.12) we have

$$R=\int_{A_x} dx \int_{A_\gamma} (\gamma-x)^2 P(x,\gamma)\,d\gamma=\overline{(\gamma-x)^2}, \tag{17.24}$$

i. e., for a quadratic loss function, average risk R coincides with mean square $\overline{(\gamma-x)^2}$ of the message reproduction error and, consequently, the minimum average risk criterion coincides with the minimum root-mean-square error criterion.

As will be demonstrated below, in many essentially-interesting cases, the minimum average risk criterion coincides also with the maximum inverse probability criterion (or maximum inverse probability density criterion).

Consequently, selecting some type of loss function $l(x,\gamma)$, we may obtain different optimizations from (17.12), including all the simpler criteria examined in preceding chapters as well.

For practical use, expression (17.12) should be transformed so that given /314 a priori probabilities P(x) and $P_x(y)$ and decision rule $\Delta(\gamma|y)$ are included in it in an explicit form. We will perform this transform, assuming that the decision rule is regular, i. e., is described by regular dependence (17.7).

Considering that

$$P(x,\gamma)=P(x)P_x(\gamma),$$

from (17.12) we obtain

$$R=\int_{A_x} dx\, P(x) \int_{A_\gamma} l(x,\gamma) P_x(\gamma) d\gamma. \tag{17.25}$$

Having designated

$$R_x=\int_{A_\gamma} l(x,\gamma) P_x(\gamma)\,d\gamma, \tag{17.26}$$

we have

$$R=\int_{A_x} P(x) R_x\,dx. \tag{17.27}$$

It follows from expressions (17.26) and (17.27) that R_x is a conditional risk determined for a given message x value, while the average risk may be obtained by averaging this conditional risk with respect to all possible message x values.

Expression (17.12) may be transformed also in the following manner. Since the following regular dependence exists between y and γ

$$\gamma = \Gamma(y),$$

then

$$P(x, \gamma)\,dx\,d\gamma = P(x, y)\,dx\,dy,$$

where

$$\gamma = \Gamma(y),$$

and; from (17.12) we have

$$R = \int_{\lambda_y} \int_{\lambda_x} P(x, y)\, l\,[x, \Gamma(y)]\,dx\,dy. \tag{17.28}$$

[Since integration bounds with respect to y will not depend on x and vice versa, then any integration procedure may be used with respect to x and y in expression (17.28)].

Expression (17.28) is convenient because given a priori distribution P(x, y) and decision rule Γ(y) are included in it in explicit form.

The optimum decision rule Γ(y) is the one in which expression (17.28) is minimal, i. e., in which

$$R = \int_{\lambda_y} \int_{\lambda_x} P(x, y)\, l\,[x, \Gamma(y)]\,dx\,dy = \min. \tag{17.29}$$

Since

$$P(x, y) = P(y)\, P_y(x),$$

then it also is possible to represent expression (17.28) in the following form: /315

$$R = \int_{A_y} P(y)\, R_y\, dy, \tag{17.30}$$

where

$$R_y = \int_{A_x} P_y(x)\, l\,[x, \Gamma(y)]\, dx. \tag{17.31}$$

It follows from relationships (17.30) and (17.31) that R_y is a conditional risk corresponding to given realization y and average risk R may be found by averaging this conditional risk with respect to all possible realizations y.

Since P(y) and R_y are non-negative functions, then it follows from (17.30) that conditional risk R_y must be minimal for all possible realizations y in order to obtain minimum average risk R.

Consequently, that decision rule $\Gamma_{\text{нр}}(y)$ which for any y minimizes conditional risk R_y magnitude is optimum, i. e., the one which insures that this condition is met

$$R_y = \int_{A_x} P_y(x)\, l\,[x, \Gamma(y)]\, dx = \min \quad \text{for any } y. \tag{17.32}$$

Correspondingly optimum is that receiving device which, for each given realizaton y(t), produces an output "decision" with respect to the rule

$$\gamma = \Gamma_{\text{нр}}(y), \tag{17.33}$$

where $\Gamma_{\text{нр}}(y)$ -- decision rule $\Gamma(y)$ corresponding to the minimum (for any y) conditional risk R_y.

We will compare this decision rule with the simpler rule of maximum inverse probability density $P_y(x)$ examined in preceding chapters. Here, for brevity,

we will call the rule based on maximum inverse probability density $P_y(x)$ a simple rule, and the rule minimizing conditional risk R_y for all y and, consequently, average risk R as well, a generalized rule.

When the simple rule is used, the decision γ selected is that message $x_{y\,M}$ value which, for any y, is converted into the inverse probability density $P_y(x)$ maximum, i. e., the simple rule has the form:

$$\gamma = x_{y\,M}. \tag{17.34}$$

where x_{yM} -- message x value at which this condition is met

$$P_y(x) = \max \quad \text{for any y.} \tag{17.35}$$

The generalized rule has the (17.33) form and is converted to the minimum (for any y) conditional risk R_y, i. e., will be found from condition (17.32).

It follows from a comparison of expressions (17.32) and (17.35) that a generalized decision rule in the general case is more complex for computations and for practical realization since expression (17.32) includes loss function /316 $l\,[x,\ \Gamma(y)]$. as well as inverse probability density $P_y(x)$. However, introduction of the loss function makes it possible to consider the relative danger of different errors as decision γ is developed.

Thus, for example, given simple reproduction and a loss function of the type

$$l\,(x,\gamma) = (\gamma - x)^2,$$

expression (17.32) takes the form

$$R_y = \int_{\Lambda_x} P_y(x)\,[x - \Gamma(y)]^2\,dx = \min \quad \text{for any } y. \tag{17.36}$$

Here, the greater the error $|x - \gamma|$, modulus, the more weight it has in expression (17.36). Therefore, generalized decision rule $\Gamma_{HP}(y)$, found from condition (17.36) considers that large errors are more dangerous since this circumstance is not

considered at all in a simple decision rule. However, for several cases, results provided by simple and generalized decision rules coincide.

For instance, let the loss function have the following special form:

$$l(x, \gamma) = c - \delta(x - \gamma), \tag{17.36a}$$

where c -- some constant, while $\delta(x - \gamma) - \delta$-delta-function. Substituting function (17.36a), called a simple loss function, into expression (17.31), we obtain

$$R_y = c - P_y(x)|_{x=\Gamma(y)}.$$

Consequently, the minimum (with respect to γ) magnitude of conditional risk R_y and, consequently, of average risk R as well, will occur if the decision $\gamma = \Gamma(y)$ selected coincides with that x value corresponding to the inverse probability density $P_y(x)$ maximum.

Thus, for a simple loss function, a simple decision rule (based on the maximum inverse probability density criterion) provides the same result as a generalized decision rule (based on the minimum average risk criterion).

As will be shown in the next section, when certain additional conditions often occurring in practical cases are met, results obtained from simple and generalized decision rules coincide, not only for a (17.36a)-type simple loss function, but in a broad class of loss functions finding wider use as well.

17.3 Minimax Optimization

The minimum average risk criterion is based upon use of complete information on the laws of signal and noise distribution.

On the one hand, this is an advantage of the criterion since it makes it possible to use the maximum-possible information about the signal and noise /317 when reproducing the message. However, this very special feature of the criterion causes difficulties in those cases when, according to problem conditions, signal and noise distributions are not precisely known. A priori distributions of messages

P(x) quite often turn out to be unknown. In some cases, even which magnitude (which time function) is the needed parameter--random or regular, may be unknown. Different approaches to surmounting this so-called "a priori difficulty" are examined in Chapter 21. We will dwell now only on the following two methods.

1. Uniform distribution P(x) method.

If distribution P(x) is completely unknown or it even is unknown whether message x is random, then it is most natural to assume in this case that distribution P(x) is uniform, i. e., to consider

$$P(x) = k, \tag{17.37}$$

where k -- some constant.

Stemming from distribution P(x) selected in this manner, it is possible further to determine average risk R in the normal manner, from formula (17.25) for example. Since it is necessary when seeking the optimum decision rule to find only minimum risk R, the magnitude of constant k plays no role here and it may be random.

The drawback to this method is that if, in actuality, distribution P(x) will turn out to be very irregular, decision rule $\Gamma(y)$ found may turn out to be far from optimum.

2. Minimax optimization criterion.

This criterion will reduce the maximum-possible value of conditional risk R_x, rather than average risk R, to the minimum.*

Mathematically, the minimax criterion is formulated in the following manner:

$$\max_{(\text{no } x)} R_x(\Delta_M) \leqslant \max_{(\text{no } x)} R_x(\Delta) \quad \text{for all } \Delta, \tag{17.38}$$

where $\Delta = \Gamma(y)$ -- random decision rule, while $\Delta_M = \Gamma_M(y)$ -- decision rule insuring

*Hence the term "minimax"--"MINImization of the MAXimum."

minimization of the maximum conditional risk R_x value and therefore called a <u>minimax decision rule</u>.

The symbol $R_x(\Delta)$ underscores that conditional risk R_x will depend on decision rule Δ.

It follows from determination of the minimax criterion that it provides the best decision rule (i. e., best receiving device operating principle) for the worst case. This is the essence of its advantage, but it is linked to its shortcoming: since, in actuality, the worst case may not occur at all or be of low probability, then a minimax decision rule may turn out (in all cases or in a majority of cases) to be far from optimum.

Wald, stemming from several sufficiently-general prerequisites, obtained /318
the following important results for the minimax criterion [30, 16, and 118]:

1. A minimax decision rule exists.

2. Any minimax decision rule Δ_M is a Bayes decision rule determined for some a priori distribution $R_M(x)$, called the least-favorable a priori distribution.

3. A least-favorable a priori distribution exists.

4. The Bayes decision rule $\Delta_{кр}$ [i. e., $\Gamma_{кр}(y)$], in which conditional risk R_x will not depend on x is a minimax decision rule (Δ_M).

5. The minimum (with respect to all rules Δ) value of the maximum [with respect to all distribution P(x) types] average risk corresponds to a minimax decision rule.

We will recall that the term Bayes is understood to mean that decision rule which minimizes average loss, in particular, average risk R.

It follows from the aforementioned results, first, that satisfaction of relatively-general conditions usually occurring in practice suffices for existence of a minimax decision rule.

Second, minimax decision rule Δ_M may be considered a Bayes decision rule insuring receipt of minimum average risk for some least-favorable a priori message $P_M(x)$ distribution.

In short, a minimax criterion insures receipt of minimum average risk for the least-favorable a priori message distribution. Therefore, if it turns out to be possible to compute the least-favorable a priori distribution $P_M(x)$, then the optimum decision rule will be found next using the conventional method examined in the next section.

In the general case, computation of distribution $P_M(x)$ presents great difficulties. However, in some cases, it will be found relatively simply through use of result 4 listed above, which means that Bayes decision rule $\Gamma_{\text{кр}}(y)$, in which conditional risk R_x will not depend on x, is minimax rule $\Gamma_M(y)$.

The methodology for finding distribution $P_M(x)$ here may comprise the following:

1. We write the expression for average risk R, in the (17.25) form, for example:

$$R = \int_{A_x} dx\, P(x) \int_{A_\gamma} l(x, \gamma)\, P_x(\gamma)\, d\gamma. \tag{17.25}$$

2. We will find Bayes decision rule $\Gamma_{\text{кр}}(y)$, i. e., the decision rule in which risk R is minimal. It follows from (17.25) that this decision rule will depend on the type of a priori P(x) distribution selected during computation of the expression, i. e.,

$$\Gamma_{\text{кр}}(y) = \Gamma_{\text{кр}}[y, P(x)]. \tag{17.39}$$

[During computation of expression (17.25), distribution P(x) is written in /319 general form since it still is unknown to us].

3. We substitute decision rule $\Gamma_{\text{кр}}(y)$ found into an expression for conditional risk R_x, for instance into expression (17.26), i. e., we assume

$$
\left.\begin{array}{c}
R_x = \int\limits_{A_\gamma} l(x, \gamma) P_x(\gamma)\, d\gamma, \\
\text{where} \\
\gamma = \Gamma_{\text{кр}}(y) = \Gamma_{\text{кр}}[y, P(x)].
\end{array}\right\} \tag{17.40}
$$

Here, conditional risk R_x turns out to be dependent on a priori distribution $P(x)$.

4. We try to select the distribution $P(x)$ type in such a way that conditional R_x determined from expression (17.40) will not depend on x. If this is done successfully, then, as follows from what was presented above, the a priori distribution found in this manner is needed distribution $P_M(x)$ also.

5. Having substituted the distribution $P_M(x)$ found into expression (17.39), we will find the minimax decision rule:

$$
\Gamma_{\text{м}}(y) = \Gamma_{\text{кр}}[y, P_{\text{м}}(x)]. \tag{17.41}
$$

Thus, in those cases when distribution $P(x)$ is selected successfully so that conditional risk R_x determined from expression (17.40) will not depend on x, in so doing least-favorable a priori distribution $P_M(x)$ and minimax decision rule $\Gamma_M(y)$ also will be found.

If it turns out that conditional risk R_x determined from expression (17.40) will depend on x regardless of the form of distribution $P(x)$, then this still does not signify that least-favorable distribution $P_M(x)$ does not exist. It only denotes that it is impossible, using the given method, to find distribution $P_M(x)$ in the case under examination and it must be sought using other methods.

Using the aforementioned methodology, Middleton demonstrated, in particular, that least-favorable distribution $P_M(x)$ is a uniform distribution, given independent additive signal and noise and simple reproduction of the shape of a signal* with a quadratic loss fur- -:· [of the (17.9) type][118].

*That is, when $x = x(t) = u_c(t)$ (See the next section).

Consequently, in this case important in practice, the following uniform a priori distribution is the least favorable

$$P(x) = \text{const},$$

and minimax decision rule $\Gamma_M(y)$ coincides with Bayes decision rule $\Gamma_{\text{кр}}(y)$ found for the uniform a priori message distribution.

17.4 Some General Results in Use of the Theory of Statistical Decisions for Analog Message Reception /320

When individual analog message values are received, the most-widespread loss function is a quadratic function of the type

$$l(x, \gamma) = (x - \gamma)^2, \tag{17.42}$$

and optimum decision rule $\Gamma_{\text{кр}}(y)$ may be found from condition (17.36), which, following evident transforms, will lead to the following form:

$$\begin{aligned} R_y = \int_{A_x} x^2 P_y(x)\,dx - 2\Gamma(y) \int_{A_x} x P_y(x)\,dx + \\ + \Gamma^2(y) \int_{A_x} P_y(x)\,dx = \min. \end{aligned} \tag{17.43}$$

Consequently, the optimum value of rule $\Gamma(y)$ is the one in which expression (17.43) is minimal, i. e., derivative $dR_y/d\Gamma$ equals zero:

$$\frac{dR_y}{d\Gamma} = -2 \int_{A_x} x P_y(x)\,dx + 2\Gamma(y) \int_{A_x} P_y(x)\,dx = 0. \tag{17.44}$$

Considering normality conditions (4.7) and (17.44), the result is

$$\Gamma_{\text{кр}}(y) = \int_{A_x} x P_y(x)\,dx. \tag{17.45}$$

It follows from (17.45) that, in this case, optimum decision rule $\Gamma_{\text{кр}}(y)$ equals the conditional (i. e., for a given y) expected message x value.

Since distribution $P_y(x)$ in the case of reception of individual analog message

values is a unidimensional probability density, then formula (17.45) has a simple geometric interpretation. $\Gamma_{\text{нр}}(y)$ equals the X-axis of the "center of gravity" of the area located under curve $P_y(x)$. Consequently, in this case, the optimum receiver, during reception of given realization y(t), must select as decision the X-axis of the "center of gravity" of the area located under curve $P_y(x)$. Here, it will insure the minimum average risk, i. e., in this case (for a quadratic loss function) the minimum root-mean-square error.

Known (given) a priori distributions P(x) and $P_x(y)$ should be used for practical realization of algorithm (17.45) to express the a posteriori (inverse) probability density $P_y(x)$ included in the logarithm. To do so, we will use relationship (4.5). Substituting it into (17.45), we obtain

$$\Gamma_{\text{нр}}(y) = \frac{\int_{\Lambda_x} xP(x)P_x(y)\,dx}{P(y)}.$$

But /321

$$P(y) = \int_{\Lambda_x} P(x,y)\,dx = \int_{\Lambda} P(x)P_x(y)\,dx,$$

therefore, finally we have

$$\gamma = \Gamma_{\text{нр}}(y) = \frac{\int_{\Lambda_x} xP(x)P_x(y)\,dx}{\int_{\Lambda_x} P(x)P_x(y)\,dx}. \tag{17.46}$$

Since a priori distributions P(x) and $P_x(y)$ are assumed known, expression (17.46) makes it possible in principle to compute optimum decision rule $\Gamma_{\text{нр}}(y)$, i. e., to find the optimum receiver structure.

Substituting decision rule $\Gamma(y) = \Gamma_{\text{нр}}(y)$ found into expression (17.28) and considering that $P(x, y) = P(x)P_x(y)$, it then is possible also to compute the minimum value of the average risk $R = R_{\text{мин}}$ corresponding to this rule.

We now will examine reproduction of messages that are random <u>time functions</u> (rather than random <u>magnitudes</u>). Reproduction of messages that are random time

functions at the present time is called filtration. Therefore, in future, we also will use this term.

In filtration, it is necessary to differentiate between real-time and delay message reproduction. Figure 17.2 illustrates the difference between the two.

Initially, we will examine real-time reproduction. Here, for simplicity, we will assume that the time change is discrete, with digitization interval Δt

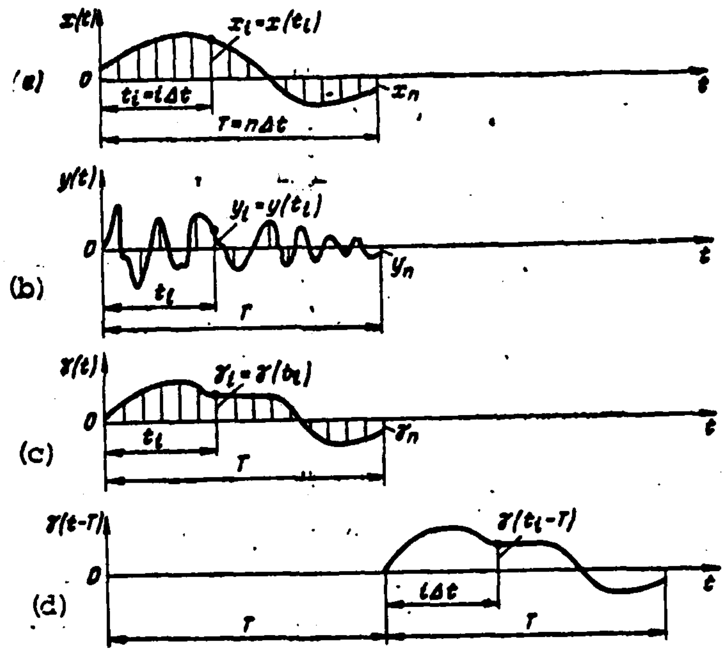


Figure 17.2

(Figure 17.2). Then, the complete duration of the reproduced message equals $T = n\Delta t$, while real time $t_i = i\Delta t$. During real-time reproduction, message $x(t)$ must be reproduced without a delay relative to real-time information $y(t)$ arriving at receiver input. This signifies that, by moment $t_1 = 1\Delta t$, it is necessary to supply decision γ_1 concerning the value of message $x_1 = x(t_1)$, while by random moment $t_i = I\Delta t$ $(i = 1, 2, \ldots, n)$, to supply decision γ_i concerning the value of message $x_i = x(i\Delta t)$ (Figure 17.2c).

In delay reproduction, a lag in shaping output realization $\gamma(t)$ for a time equal to (or exceeding) T is permissible (Figure 17.2d).

In delay reproduction, high realization x(t) fidelity may be obtained. Actually, in this case, by the moment the decision is made about any message x(t) values $(x_1, \ldots, x_i, \ldots, x_n)$, the receiver possesses entire realization y(t) during interval (0 – T). In a case of real-time reproduction, decision γ_i concerning message x_i value must be taken when only that part of realization y(t) which occurred during interval 0 – iΔt is present. Therefore, all other conditions being equal, initial realization x(t) sections (corresponding to slight values of number i) will be reproduced less accurately and only finite value $x_n = x(n\Delta t)$ will be reproduced with the same accuracy. /322

However, in a number of practical problems, a significant lag is not permitted; in addition, in delay reproduction, the algorithm of the optimum process may be complicated.

In future, we will limit ourselves to examination only of real-time reproduction. In this case, as noted above, at each moment in time $t = t_i$, the requirement is best reproduction of the message $x_i = x(t_i)$ value corresponding to a given moment in time and, consequently, the loss function must be written in the following form:

$$\left.\begin{aligned} l(x, \gamma) &= l(x_i, \gamma_i), \\ x_i = x(t_i), &\quad \gamma_i = \gamma(t_i). \end{aligned}\right\} \qquad (17.47)$$

where

Here, in accordance with (17.32), optimum decision rule $\Gamma_{\text{кр}}(y)$ is the one which insures that this condition is met

$$R_y = \int_{A_x} P_y(x_i)\, l[x_i, \Gamma(y)]\, dx_i = \min, \qquad (17.48)$$

where /323

$$y = (y_1, \ldots, y_i).$$

But

$$\tilde{P}_y(x_i) = \int_{A_{x_1}} \dots \int_{A_{x_{i-1}}} P_y(x_1, \dots, x_i)\, dx_1 \dots dx_{i-1},$$

and condition (17.48) takes the following form:

$$R_y = \int_{A_{x_1}} \dots \int_{A_{x_i}} P_y(x_1, \dots, x_i)\, l\,[x_i, \Gamma(y)]\, dx_1 \dots dx_i = \min. \qquad (17.49)$$

In future, we will limit ourselves to examination of a quadratic loss function. Then, in accordance with (17.47), one should assume

$$l(x, \gamma) = (x_i - \gamma_i)^2$$

and, from (17.49), we obtain

$$R_y = \int_{A_{x_1}} \dots \int_{A_{x_i}} P_y(x_1, \dots, x_i)\, [x_i - \Gamma(y)]^2\, dx_1 \dots dx_i = \min, \qquad (17.50)$$

i. e.,

$$\frac{dR_y}{d\Gamma} = 0.$$

Doing calculations analogous to those presented above [during derivation of formula (17.46)], we obtain

$$\left.\gamma_i = \Gamma_{\text{кр}}(y) = \frac{\int_{A_{x_1}} \dots \int_{A_{x_i}} x_i\, P(x_1, \dots, x_i)\, P_{x_1, \dots, x_i}(y)\, dx_1 \dots dx_i}{\int_{A_{x_1}} \dots \int_{A_{x_i}} P(x_1, \dots, x_i)\, P_{x_1, \dots, x_i}(y)\, dx_1 \dots dx_i},\right\} \qquad (17.51)$$

where $y = (y_1, \dots, y_i)$, $i = 1, 2, \dots, n$.

Since $n \gg 1$, then essentially it is impossible to realize such an algorithm for optimum receiver action due to its extreme complexity. Algorithms applicable for practical use have been obtained only in those cases when it turned out to be possible to represent multidimensional distributions $P(x_1, \dots, x_n)$ and $P_{x1, \dots, x_n}(y_1, \dots, y_n)$ in a simple-enough form.

The following are the most important of these cases.

First case. Message x(t) and noise $u_{ш}(t)$ are additive, independent, and have a normal distribution, i. e.,

$$\left.\begin{array}{l} y(t)=u_c(t)+u_{ш}(t), \\ \text{where } x(t) \text{ and } u_{ш}(t) \text{ -- independent normal random processes} \end{array}\right\} \tag{17.52}$$

Evidently, in this case, signal-plus-noise y(t) also has normal distribution.

For normal random process u(t), n-dimensional distribution $P(u_1, \ldots, u_n)$ /324 may be represented in the form

$$P(u_1, \ldots, u_n)=c\exp\left[-\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{V_{ij}}{\mathfrak{D}}(u_i-\bar{u}_i)(u_j-\bar{u}_j)\right], \tag{17.53}$$

where c -- constant determined from the normality condition and, in the case under examination, not having a value [since it cancels itself after the corresponding distributions are substituted into formula (17.51)]; $\mathfrak{D}$ -- determinant of the type

$$\mathfrak{D}=|R_{ij}|, \tag{17.54}$$

whose elements are values of correlation function

$$R_{ij}=\overline{(u_i-\bar{u}_i)(u_j-\bar{u}_j)};$$

here

$$u_i=u(i\Delta t);\quad u_j=u(j\Delta t);\quad V_{ij}=(-1)^{i+j}\Delta_{ij},$$

Δ_{ij} -- minor of determinant $\mathfrak{D}$ relative to element R_{ij}.

Since processes x(t) and y(t) in the case under examination have normal distributions, (17.53)-type expressions are valid for them.

Substitution of expressions of this type into relationship (17.51) and examination of formulas (17.49) and (17.51) demonstrate that formation of output effect $\gamma(t)$ for given input realization y(t) requires that only linear operations be

performed on this realization. This will lead to a conclusion very important for practice: if the message and noise are additive, independent, and have normal distribution, then, given a quadratic loss function, the optimum message reproduction system is a linear system. In other words, for the aforementioned conditions, introduction of any distortions into the system may not decrease the magnitude of the average risk (in this case, the magnitude of the root-mean-square message reproduction error).

As analysis demonstrates, this result obtained above for a quadratic loss function retains its force also for several other loss function types and may be formulated in the following, more-general, form: if loss function $l(x, \gamma)$ will depend only on the real-time value of error $(x - \gamma)$, while signal-plus-noise has the (17.52) form, then the optimum system (providing minimum average risk) is a linear system, while the structure of this system is just like that of a quadratic loss function. Only systematic bias (expected value) of the error in this case will depend on the specific loss function type; if function $l(x, \gamma)$ is a nondecreasing function of the error $|x - \gamma|$ modulus, then systematic bias will not depend on specific loss function type either.

Thus, the first case is characterized by the fact that, for a relatively-broad class of loss functions (optimizations), optimum filtration boils down to /325 linear filtration, examined in § 2.2.

Second case. Signal-plus-noise y(t) has the form

$$y(t) = u_c(t, x) + u_m(t),$$

where $u_c(t, x)$ -- precisely-known signal [with the exception of message x = x(t), which is its parameter]; $u_m(t)$ - stationary normal white noise.

Message x(t) -- stationary random process with power spectrum close in shape to a rectangle; message fidelity is high (high signal-to-noise ratio).

In this formulation, the filtration problem was solved for the first time by Kotel'nikov in 1947 and the basic results of the solution were presented in Chapter 7. The second case is more complex than the first because that reproduced

message x(t) included in signal-plus-noise y(t) is not additive, but is in the form of a signal $u_c(t, x)$ parameter. Kotel'nikov for problem solution was not required to provide any specific type of message x(t) distribution $P(x_1, \ldots, x_n)$ since, in accordance with his assumptions, the type of this distribution did not exert a major impact on optimum receiver structure and properties (the validity of this assertion flows from Chapter 7 results and examinations presented in Chapter 19). Distribution $P_{x1, \ldots, xn}(y_1, \ldots, y_n)$, for the assumptions made about signal-plus-noise y(t) type, turns out to be normal.

Third case. Signal-plus-noise y(t) has the form

$$y(t) = u_c(t; x; \alpha_1, \ldots, \alpha_m) + u_m(t), \tag{17.55}$$

where x = x(t); $\alpha_i = \alpha_i(t)\ (i = 1, \ldots, m)$ -- parasitic signal parameters.

Here, signal-plus-noise y(t) type is more complicated than in the preceding cases and the optimum filtration system in the general case turns out to be nonlinear. Therefore, the synthesis problem solved in this case is called an optimum nonlinear filtration problem.

The starting point for solution usually is the minimum root-mean-square error criterion, i. e., relationship (17.51). Here, one of the following types of distribution law approximation is used:

1. Markov approximation.

2. Gauss approximation.

R. L. Stratonovich in 1959--1960 for the first time developed the Markov approximation method [138, 154, 155, 158, and others] and it was based on the assumption that message x(t) and signal-plus-noise y(t) are components of a multidimensional Markov process.

I. A. Bol'shakov and V. G. Repin in 1961 [156] proposed the Gauss approximation method and it was developed in subsequent works [127, 133, and others].

The following chapter is devoted wholly to the theory of optimal nonlinear filtration in view of its great importance. Here, the presentation for brevity only will cover the Gauss approximation. This is because, for a Gauss approximation, results are simpler and clearer. However, a Gauss approximation requires /326 several serious assumptions, which in principle may be avoided in the Markov approximation. Therefore, the Markov theory also has great theoretical and practical significance. Its foundations are presented briefly in [133], while its applications are examined in several works [130, 158, and others].

To conclude this chapter, we will note the following, more-particular but important, results obtained in the theory of optimum filtration. They will apply to a case when message x(t) and noise $u_m(t)$ are additive and independent and they comprise the following:

1. If a priori message distribution P(x) is uniform, then Bayes decision rule $\Gamma_{кр}(y)$ has the so-called property of translation, i. e.,

$$\Gamma_{кр}(y+\lambda)=\Gamma_{кр}(y)+\lambda, \tag{17.56}$$

where λ -- random fixed magnitude (or time function). This denotes, for example, that, if fixed message x(t) is supplied to optimum filtration system input and this message changes by magnitude λ(t), then the oscillation at system output changes by the same magnitude.

2. If no restrictions are placed on distribution P(x), then its least-favorable form, $P_M(x)$, is a uniform distribution (this special feature already was noted in the preceding section).

CHAPTER EIGHTEEN /326

OPTIMUM NONLINEAR FILTRATION

18.1 Basic Relationships During Optimum Nonlinear Filtration

As noted in § 17.4, decision rule (system operation algorithm) (17.51), which insures minimum root-mean-square error in simple message x(t) reproduction, corresponds to optimum linear filtration.* In accordance with the Gauss approximation method, the following basic assumptions are made to bring this decision rule to a form suitable for practical use:

1. Message x(t) has a normal law of distribution [with known mean value $\overline{x(t)}$ and correlation function $R_x(t_1, t_2)$].

2. Message x(t) changes slowly compared with additive noise $u_ш(t)$ and /327 with parasitic signal parameters which change over time, i. e., these conditions are met

$$\tau_{кор\,x} \gg \tau_{кор\,ш}; \quad \tau_{кор\,x} \gg \tau_{кор\,\alpha}, \tag{18.1}$$

*Some generalization to a case of complex message reproduction is presented in § 18.4.

where $\tau_{кор\,x}$, $\tau_{кор\,ш}$ and $\tau_{кор\,\alpha}$ -- message correlation time, additive noise time, and parasitic signal parameter time, respectively (signal fluctuations).

3. Message x(t) reproduction is very accurate, i. e., mean-square error

$$\overline{\varepsilon^2(t)} = \overline{[x(t)-\gamma(t)]^2} \tag{18.2}$$

is a very slight magnitude.

4. Modulation of signal $u_s(t;\ x;\ \alpha_1,\ \ldots,\ \alpha_m)$ by message x(t) -- direct (determination of direct modulation was provided in § 7.1).

5. Statistical characteristics of parasitic signal parameters $\alpha_1,\ \ldots,\ \alpha_m$ will not depend on time.

Considering these assumptions*, analysis of initial expression (17.51) provides the following results after very unwieldy transforms (see [127], for example):

1. Optimum estimate γ(t) of the reproduced message equals

$$\gamma(t) = \int_{t_0}^{t} \eta(t,\ \tau)\, z(\tau)\, d\tau + \overline{x(t)}, \tag{18.3}$$

where impulse response η(t, τ) is determined from integral equation

$$k\int_{\tau}^{t} C(t,\ s)\,\eta(s,\ \tau)\, ds + C(t,\ \tau) = \eta(t,\ \tau). \tag{18.4}$$

Additional function C(t, τ) included in this expression is in turn determined from the following integral equation:

$$k\int_{t_0}^{t} C(t,\ s)\, R_x(s,\ \tau)\, ds + C(t,\ \tau) = R_x(t,\ \tau). \tag{18.5}$$

In these formulas, t_0 -- moment that message x(t) reproduction (measurement) begins, while t -- actual moment in time.

*The validity of these assumptions is discussed in § 18.6.

2. Function z(t) and constant factor k included in the formulas are determined from the following expressions:

$$z(t) = -\left[\frac{\partial Q(y, x, t)}{\partial x}\right]_{x=\gamma}; \tag{18.6}$$

$$k = \overline{\left[\frac{\partial^2 Q(y, x, t)}{\partial x^2}\right]} \tag{18.7}$$

the superimposed dotted line denotes averaging over time for an infinitely-long /328 time interval).

Function Q(y, x, t) will be found from the equation

$$\int\limits_{t-\Delta t}^{t} Q(y, x, t)\,dt = -\ln P_x(y), \tag{18.8}$$

where $P_x(y)$ — likelihood function for realizations y(t) determined in interval $(t - \Delta t) \div t$, which may be random within the following range:

$$\tau_{\text{кор ш}},\ \tau_{\text{кор }\alpha} \ll \Delta t \ll \tau_{\text{кор м}}, \tag{18.9}$$

i. e., interval Δt must be much greater than noise correlation intervals $\tau_{\text{кор ш}}$ and $\tau_{\text{кор }\alpha}$, but less than message correlation time [in view of the assumptions made above, simultaneous satisfaction of inequality (18.9) is possible].

Function z(t) may also be represented in the following form:

$$z(t) = k\,[\varepsilon(t) + \Delta e_n(t)], \tag{18.10}$$

where, just as before,

$$\varepsilon(t) = x(t) - \gamma(t), \tag{18.11}$$

while $\Delta e_n(t)$ — stationary white noise with power spectrum (unilateral)

$$g_{ш0} = \frac{2}{k}. \tag{18.12}$$

4. Error ϵ(t) has normal distribution with zero mean value $\overline{\epsilon(t)} = 0$ and variance

$$\sigma_\epsilon^2(t) = \overline{\epsilon^2(t)} = C(t, t), \tag{18.13}$$

where function C(t, t) is obtained from additional function C(t, τ) through replacement of τ by t.

The equivalent circuit depicted in Figure 18.1 corresponds to relationships (18.3)--(18.8), (18.10), and (18.11). In this circuit, ДИС (discriminator) denotes

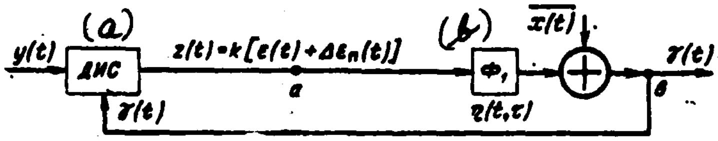


Figure 18.1. (a) -- DIS [Discriminator];
(b) -- F_1 [Filter].

the element (in the general case, nonlinear) which computes function z(t) by processing received oscillation y(t) considering measured value γ(t); Φ_1 -- linear filter with impulse response η(t, τ) determined by equations (18.4) and (18.5). It is easy to become convinced that part of the Figure 18.1 circuit connected between points a and b actually converts z(t) into γ(t) in accordance with /329 equation (18.3). As follows from Figure 18.1, the discriminator converts input signal-plus-noise y(t) into z(t). The law of this conversion is determined from relationships (18.6) and (18.8). Consequently, these relationships completely determine discriminator structure.

It follows from formula (18.6) that, when z(t) is formed from y(t), measured signal value γ(t) must be considered. Consequently, along with y(t), measured message value γ(t) must be introduced into the discriminator, i. e., a feedback circuit must exist from meter output to discriminator, as depicted in Figure 18.1. It follows from expressions (18.10)--(18.12) that discriminator ДИС supplies an

oscillation proportional to the sum of error signal $\epsilon(t)$ and white noise with spectrum 2/k, i. e., it will reproduce error signal $\epsilon(t)$ with an error characterized by noise $\Delta\epsilon_n(t)$.

It follows from Figure 18.1 and formula (18.10) that it is possible to use the equivalent circuit depicted in Figure 18.2 rather than the Figure 18.1 circuit

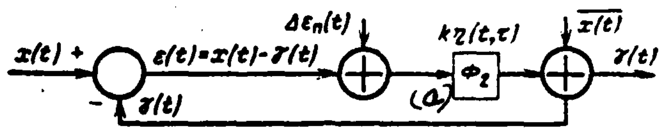


Figure 18.2. (a) -- F_2 [Filter].

when computing $\gamma(t)$. In Figure 18.2, Φ_2 -- linear filter with impulse response $k\eta(t, \tau)$. Consequently, filter Φ_2 diffe-- from filter Φ_1 (Figure 18.1) only in that its transfer constant is increased by a factor of k.

It follows from Figure 18.1 and formulas (18.10) and (18.12) that the quality of discriminator operation completely is characterized by single parameter k, which, first, is the discriminator transfer constant (rate of change) for slight signal error values and, second, determines on a one-to-one basis noise $\Delta\epsilon_n(t)$ spectral density at its output.

It also follows from examination of Figures 18.1 and 18.2 that the discriminator and linear filter Φ_1 in an optimum system perform the following functions:

a) the discriminator must separate error signal $\epsilon(t)$ from signal-plus-noise y(t) with minimum error, i. e., insure minimum noise spectral density g_{m0} at its output;

b) linear filter Φ_2 (and, therefore, filter Φ_1) must insure the minimum mean square $\overline{\epsilon^2}$ of message x(t) reproduction error in the Figure 18.2 circuit, i. e., under conditions when the sum of error signal $\epsilon(t)$ and white noise $\Delta\epsilon_n(t)$ with spectral density equalling 2/k arrives at its input.

Independent discriminator and linear filter system design is possible. Actually,

the optimum discriminator structure completely is determined by expressions (18.6) and (18.8), which will not depend directly on filter Φ_1 parameters. It is /330 possible during optimum filter Φ_1 (or Φ_2) system design to start from equations (18.4) and (18.5) or directly from the Figure 18.2 circuit if you assume given message correlation function $R_x(t_1, t_2)$ and require that minimum mean-square error $\epsilon(t)$ be obtained. Here, if the discriminator still is not considered and its parameter k still not determined, it is possible to obtain a signal decision in a general form, i. e., for random value k.

Since the majority of the assumptions made previously were qualitative [see relationship (18.1), for example], the following approach is possible reliably to determine to what degree a system structure found will be optimum for actual message, signal, and noise characteristics. For the system structure found, compute mean-square error $\overline{\varepsilon^2}$ through analysis of the passage of signal and noise through the system and compare the resultant value with the value computed from formula (18.13). If both values $\overline{\varepsilon^2}$ turn out to be close to each other, then this will signify that the system structure found actually is sufficiently-close to optimum.

It should be noted that, during optimum linear filter Φ_2 (and, of Φ_1 as well) system design, it is possible in the Figure 18.2 circuit to assume that $\overline{x(t)} = 0$ if you simultaneously accept that the mean message x(t) value at system input also equals zero. Here, the result obtained will be valid also in the case when, in actuality, $\overline{x(t)} \neq 0$. This is explained by the fact that the mean value of measured message $\gamma(t)$ in the Figure 18.2 circuit equals $\overline{x(t)}$ and, as a result of this, during shaping of error signal $\epsilon(t)$, the mean values of processes x(t) and $\gamma(t)$ mutually compensate for each other and they do not reach filter Φ_2 input.

Consequently, during filter Φ_2 system design, it is possible to begin with the Figure 18.3a equivalent circuit. Here, the assumption is that $\overline{x(t)} = 0$. It is possible also to represent the Figure 18.3a circuit in the form depicted /331 in Figure 18.3b since the transfer of disturbance $\Delta e_n(t)$ to system input does not change the effect at its output. But, linear filter Φ_2 encompassed by feedback also is a linear system with some impulse response $\eta_{*}(t, \tau)$. Therefore, the Figure 18.3b circuit may be represented in the form shown in Figure 18.3c, where Φ_{*} — some equivalent linear filter with impulse response $\eta_{*}(t, \tau)$.

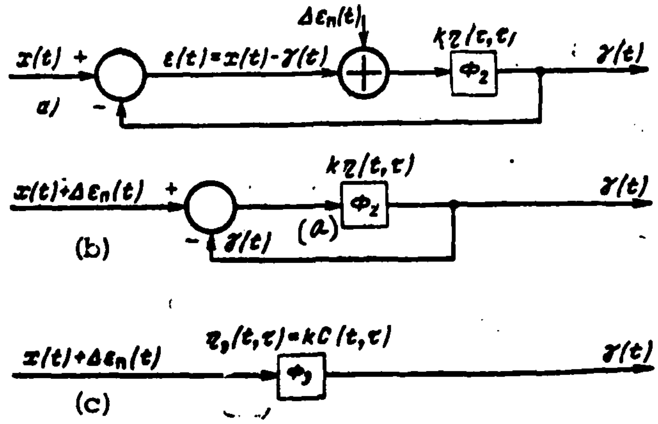


Figure 18.3

It follows from Figure 18.3c that filter Φ_3 must insure reproduction, with minimum root-mean-square error, of message x(t) arriving at its input in a mixture with additive white noise $\Delta\varepsilon_n(t)$. This problem fully coincides with that presented in § 2.2 for optimum linear filtration.

Therefore, optimum impulse response $\eta_3(t, \tau)$ must be determined by a (2.21a)-type integral equation, i. e., by the equation

$$\int_{t_0}^{t} R_x(\tau, s)\,\eta_3(t, s)\,ds + S_{m0}\eta_3(t, \tau) = R_x(t, \tau).$$

Comparing this with equation (18.5) and considering that, in the given case

$$S_{m0} = \frac{g_{m0}}{2} = \frac{1}{k},$$

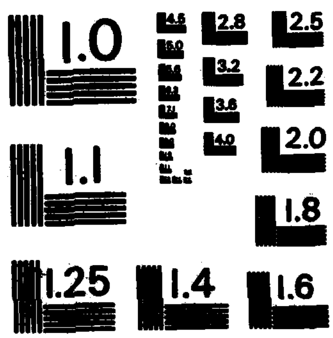
we obtain

$$\eta_3(t, \tau) = kC(t, \tau),$$

i. e., previously-introduced additional function C (t, τ) is (precise to constant factor k) the impulse response of equivalent linear filter Φ_3, accomplishing optimum filtration of message x(t) in the Figure 18.3c circuit.

AD-A120 899 THEORY OF OPTIMUM RADIO RECEPTION METHODS IN RANDOM NOISE(U) FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OH L S GUTKIN 24 SEP 82 FTD-ID(RS)T-0784-82 6/7

UNCLASSIFIED F/G 9/4 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

It follows from relationship (18.4) that impulse response $\eta(t,\tau)$ of optimum linear filter Φ_2 is linked with impulse response $C(t, \tau)$ also by an integral equation. Thus, two integral equations, (18.4) and (18.5), which determine the structure of optimum linear filter Φ_1 included in the Figure 18.1 system, have the following meaning: a) equation (18.5) makes it possible to find impulse response $kC(t,\tau)$ of the equivalent linear filter connected as in the Figure 18.3c circuit from the characteristics of message x(t) and white noise $\Delta e_{\text{п}}(t)$; b) equation (18.4) makes it possible from impulse response $kC(t,\tau)$ found for this filter to determine the corresponding impulse response of linear filter Φ_2 (or Φ_1), i. e., of the filter encompassed by feedback. Thus, optimum linear filter Φ_1 structure is determined by two, rather than one, integral equations because this filter is encompassed by feedback.

Hence, it also follows that the theory of optimum nonlinear filtration includes the theory of optimum linear filtration as a particular case.

It follows from relationships (18.6)--(18.9) that, during optimum discriminator system design, input mixture y(t) is examined in interval Δt, slight compared to $\tau_{\text{кор}\,x}$, but large compared to $\tau_{\text{кор}\,\alpha}$ and $\tau_{\text{кор}\,\text{ш}}$. This signifies that the optimum discriminator, having significant sluggishness (and, consequently, filtering /332 properties as well) with regard to noise and signal fluctuations, at the same time is essentially inertia-free with regard to reproduced message x(t).

Thus, the optimum system turns out to comprise two considerably-different elements: a nonlinear (in the general case) element inertia-free with regard to message x(t) (discriminator) and a sluggish [both for noise and for message x(t)] linear element--filter Φ_1.

This is due to the nature of the assumptions made during system design, primarily those concerning the slowness of the message x(t) change and insignificance of the error in its reproduction.

Actually, for example, let us design a radio receiving system operating in accordance with the simplest circuit, i. e., comprising an rf amplifier (including rf amplifier, frequency converter, and if amplifier), detector, and lf amplifier. It is known that, given powerful noise and a broad message spectrum, best message

reproduction requires primarily narrowing the bandwidth ahead of the detector, even if this causes some message distortions.

If this receiver is designed for high message fidelity and, consequently, will operate in noise that is not too strong, while the message spectrum width is small, then it is possible, without sacrificing message fidelity, to make the bandwidth ahead of the detector sufficient for undistorted (when there is no noise) message reproduction and to narrow the bandwidth only beyond the detector.

Analogous to this and in the optimum system examined above, it turns out to be possible, given such conditions, to make the entire system essentially inertia-free with regard to reproduced message x(t), with the exception of sluggish linear input networks--linear filter Φ_1.

The circumstance that optimum filter Φ_1 at discriminator output is a linear system is stipulated during system design by the assumption accepted concerning the normal message x(t) law of distribution. Given another law of message distribution, optimum smoothing networks at discriminator output turn out to be nonlinear.

However, the formulas presented above for optimum system design are valid also for a law of message distribution differing from the norm if, in accordance with problem conditions (with respect to considerations of simplicity, for example), filter Φ_1 at discriminator output must be linear.

We now will explain the significance of the feedback with respect to /333
reproduced message x(t) present in the optimum system (Figure 18.1). When error

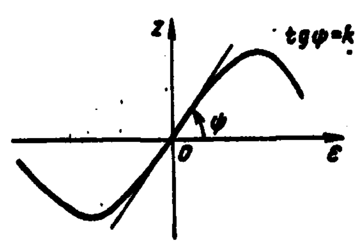


Figure 18.4

signal $\epsilon(t)$ values are slight, discriminator rate of change equals constant magnitude k, i. e., discriminator response curve $z = z(\epsilon)$ is linear. However, when ϵ values are large, this response curve turns out to be nonlinear (Figure 18.4). Therefore, difference $\varepsilon = x - \gamma$ must be kept sufficiently small in order for the discriminator to operate normally. The true message x(t) value will be contained in mixture y(t) arriving at discriminator input; the reproduced value of this message $\gamma(t)$ is shaped only at the output of this system. Therefore, it is necessary to introduce $\gamma(t)$ into the discriminator via the feedback network, as depicted in Figure 18.1.

The necessity to introduce a priori known mean message value $\overline{x(t)}$ at system output (Figure 18.1) also is quite understandable. As previously noted, thanks to this, determinate component $\overline{x(t)}$ contained in reproduced message x(t) turns out to be compensated for at the input of filter Φ_2, and of Φ_1 also, and, consequently, filter Φ_1 no longer must reproduce it. This makes it possible, all other conditions being equal, to narrow the filter Φ_1 bandwidth and to decrease the noise action accordingly.

So, the aforementioned relationships make it possible to determine the structure of the optimum message x(t) reproduction system and to find the potential (minimum-possible) value of message x(t) reproduction error $\sqrt{\overline{\varepsilon^2(t)}}$.

All these relationships were obtained for a case when signal-plus-noise y(.) at system input may be represented in the form of expression (17.55). However, the case often occurs when

$$y(t) = u_s(t;\ \varepsilon;\ \alpha_1,\ \ldots,\ \alpha_m) + u_m(t), \tag{18.14}$$

i. e., it is not reproduced parameter x(t) it[illegible]nt is introduced directly into the expression for the signal, but rather signal error

$$\varepsilon(t) = x(t) - \gamma(t); \tag{18.15}$$

the task, as usual, boils down to best message x(t) reproduction, i. e., to meeting condition

$$\overline{\varepsilon^2(t)} = \min. \tag{18.16}$$

Such a situation occurs, in particular, when designing automatic target tracking system using angular coordinates.

A functional schematic of such a system is depicted in Figure 18.5. Axis /334
z_a of this system's equisignal zone at each moment in time is directed at the

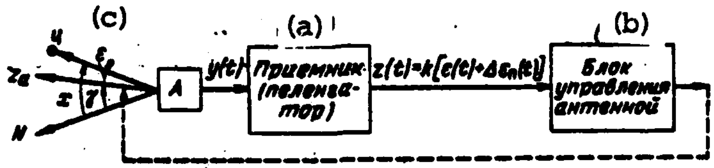


Figure 18.5. (a) -- Receiver (direction finder); (b) -- Antenna control unit; (c) -- Target.

target with the aid of motors included in the antenna control unit. Error signal z(t) characterizing deviation ϵ of axis z_a from the direction to the target reaches the input of this unit from radio receiving device output. It is possible, given slight angular deviations, to assume that

$$z(t) = k[\varepsilon(t) + \Delta\varepsilon_п(t)], \tag{18.17}$$

where k - radio receiving device (radio direction finder) gain with respect to the error signal, $\Delta\varepsilon_п(t)$ -- measurement error caused by noise action and reduced to angular deviation ε.

Evidently, in this case, the target angular coordinate and equisignal zone z_a angular coordinate (coordinates in a certain fixed system) play the role of reproduced message x(t) and the result of its reproduction γ(t) supplied by the system, respectively. Designations in Figure 18.5 for simplicity are given relative to operation of one of two antenna control system channels (for the angle of sight antenna control channel, for example); therefore, instead of vector magnitudes $\vec{x}$, $\vec{\gamma}$, and $\vec{\epsilon}$, only their scalar components located in one plane are supplied, while the fixed coordinate system is supplied by direction N.

Given highly-accurate automatic tracking, the antenna control unit may be considered linear and the autotracking system structural schematic may be represented

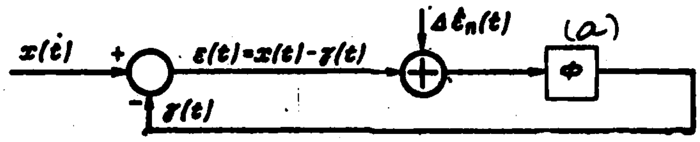


Figure 18.6. (a) -- F [Filter].

in the form shown in Figure 18.6. In this figure, Φ -- linear filter with transfer constant equalling (precise to constant factor k) the antenna control unit transfer function.

This structural schematic completely coincides with the servomechanism structural schematic depicted in Figure 18.3a. Therefore, the circumstance that, in this case, the automatic target tracking system structural schematic was given beforehand, i. e., prior to beginning mathematical system design, does not hinder obtaining optimum results corresponding to the case of the above optimum servomechanism system design.* The difference lies only in the fact that, in the case of the autotracking system under examination (Figure 18.5), mixture y(t) of /335 signal with additive noise must be represented in the (18.14) form, rather than the (17.55) form, since the signal at antenna system output is modulated by error signal $\epsilon(t) = x(t) - \gamma(t)$, rather than by message x(t) itself.

It is clear from the above optimum system analysis that, here, all relationships characterizing this system's structure and properties remain valid if, in formulas (18.6), (18.7), and (18.8), signal parameter x is replaced by new signal parameter ϵ, i. e., to assume

*Strictly speaking, the difference will be that, in the optimum system (Figure 18.1), input of a priori known expected value $\overline{x(t)}$ is envisaged, while, in the Figure 18.5 system, it is not envisaged. Therefore, if $\overline{x(t)} \neq 0$, then input of function $\overline{x(t)}$ at antenna control unit output must be envisioned to eliminate this difference in the Figure 18.5 system.

$$z(t) = -\left[\frac{\partial Q(y, \varepsilon, t)}{\partial \varepsilon}\right]_{\varepsilon=0}, \tag{18.18}$$

$$k = \overline{\left[\frac{\partial^2 Q(y, \varepsilon, t)}{\partial \varepsilon^2}\right]}, \tag{18.19}$$

$$\int_{t-\Delta t}^{t} Q(y, \varepsilon, t)\, dt = -\ln P_\varepsilon(y). \tag{18.20}$$

Here, as before, $P_\varepsilon(y)$ — likelihood function, i. e., realization y(t) (at interval Δt) probability density (multidimensional) for a given realization of signal parameter ϵ (t).

Examples of the use of the resultant general relationships to determine optimum system structure (optimum discriminator and linear filter Φ_1 structures) and computation of potential message fidelity will be presented in subsequent sections.

18.2 Optimum Discriminator System Design

The following sequence of operations for optimum discriminator system design flows from general relationships (18.6)—(18.8) or (18.18)—(18.20) and (18.12) presented in the preceding section.

First, we compute likelihood function $P_x(y)$ and its logarithm $\ln P_x(y)$. Then, the expression found for $\ln P_x(y)$ is written in the form of an integral from some time function in the (t — Δt) ÷ t range. Here, the integrand also will be Q(y, x, t).

Next, differentiating Q(y, x, t) with respect to x and replacing x for in the resultant expression, we will find oscillation z(t) at discriminator output. The resultant expression for z(t) demonstrates which mathematical operations must be performed on received oscillation y(t) in the discriminator, i. e., determines discriminator structure.

It follows from expressions (18.7) and (18.12) that the spectral density $g_{ш\varepsilon}$ of noise $\Delta\varepsilon_ш(t)$, which is reduced to error signal ϵ (t), is determined from the formula

$$g_{m\theta} = \frac{2}{k} = \frac{2}{\overline{\overline{\left[\frac{\partial^2 Q(y, x, t)}{\partial x^2}\right]}}}, \qquad (18.21)$$

i. e., computation of $g_{m\theta}$ requires finding the second derivative with respect /336 to x from Q(y, x, t) and averaging the resultant expression with respect to all possible realizations (superimposed thin line) and with respect to time (superimposed dotted line). If the process is ergodic, then averaging only with respect to time in formula (18.21) suffices.

As our example, we will examine a case when the signal is determinate (precisely known, except for x), while additive noise $u_m(t)$ has the form of normal white noise. Here, the likelihood function is determined from expression (4.10), which in this case may be written in the form

$$P_x(y) = c \exp\left\{-\frac{1}{N_0}\int_{t-\Delta t}^{t}[y(t) - u_0(t, x)]^2\, dt\right\}, \qquad (18.22)$$

where c -- some constant.

Therefore

$$-\ln P_x(y) = -\ln c + \frac{1}{N_0}\int_{t-\Delta t}^{t}[y(t) - u_0(t, x)]^2\, dt. \qquad (18.23)$$

Since only partial derivatives with respect to x will be included in the expressions for z(t) and $g_{m\theta}$, then, as is easy to see, expression (18.23) terms not depending on x do not affect the results and, consequently, they may be disregarded. Here, comparison of expressions (18.8) and (18.23) provides

$$Q(y, x, t) = \frac{1}{N_0}[y(t) - u_0(t, x)]^2. \qquad (18.24)$$

Substituting this relationship into (18.6), we obtain

$$z(t) = y(t)\,u_1(t, \gamma) + u_2(t, \gamma), \qquad (18.25)$$

where

$$u_1(t, \gamma) = \frac{2}{N_0}\left[\frac{\partial u_c(t, x)}{\partial x}\right]_{x=\gamma} \qquad (18.26)$$

$$u_2(t, \gamma) = -\frac{2}{N_0} u_c(t, \gamma)\left[\frac{\partial u_c(t, x)}{\partial x}\right]_{x=\gamma}. \qquad (18.27)$$

The discriminator structural schematic depicted in Figure 18.7 corresponds to these expressions, i. e. the discriminator will comprise a synchronous detector

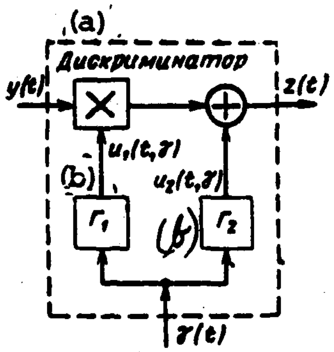


Figure 18.7. (a) -- Discriminator;
(b) -- G [Generator].

(multiplier), adder, and generators $Г_1$ and $Г_2$ of time functions $u_1(t, \gamma)$ and $u_2(t, \gamma)$, the only unknown parameter of which is oscillation $\gamma(t)$ supplied from servomechanism output (see Figure 18.1).

For example, let signal $u_c(t, x)$ be modulated with respect to amplitude by the message, i. e.,

$$u_c(t, x) = U_0[1 + \mu x(t)] B(t), \qquad (18.28)$$

where $U_0B(t)$ -- signal carrier oscillation. Then, from formulas (18.26) and (18.27), we obtain

$$u_1(t, \gamma) = \frac{2\mu}{N_0} U_0 B(t) \qquad (18.29)$$

and

$$u_2(t, \gamma) = -\frac{2\mu}{N_0} U_0^2 B^2(t)\,[1+\mu\gamma(t)]. \tag{18.30}$$

Since signal carrier oscillation $U_0B(t)$ and parameters μ and N_0 are assumed known, then these expressions completely determine generator Γ_1 and Γ_2 structure.

The following expression is obtained (considering the insignificance /337 of magnitude $\varepsilon = x - \gamma)$ from formulas (18.21) and (18.24) for the spectral density of the noise at discriminator output:

$$g_{m0} = \frac{N_0}{\overline{\left[\frac{\partial u_c\,(t,\,x)}{\partial x}\right]^2}}. \tag{18.31}$$

For amplitude modulation, when signal $u_c(t, x)$ is determined from expression (18.28), formula (18.31) provides

$$g_{m0} = \frac{N_0}{\mu^2 P_{0\,cp}}, \tag{18.32}$$

where

$$P_{0\,cp} = U_0^2\,\overline{B^2(t)} \tag{18.33}$$

-- specific carrier oscillation mean power (here and in future, power developed in a unitary resistance is called specific).

We now will examine a case of a fluctuating signal of the type

$$u_c(t;\, x;\, \alpha_1,\, \ldots,\, \alpha_m) = u(t,\, x)\,E(t)\cos[\omega_0 t + \psi(t,\, x) + \varphi(t)]. \tag{18.34}$$

Here, $E(t)$ and $\varphi(t)$ -- signal amplitude and phase fluctuations, respectively, while the assumption is that $E(t)\cos\varphi(t)$ and $E(t)\sin\varphi(t)$ have normal laws of distribution with zero mean values and known correlation function $\rho(\tau)$. Here, $E(t)$ is normalized so that

$$\overline{E^2(t)} = 2P_{cp}, \tag{18.35}$$

where

$$P_{cp} = \overline{u_c^2(t;\ x;\ \alpha_1,\ \ldots,\ \alpha_m)}$$

— specific mean signal power; u(t, x) and ψ(t, x) — usable amplitude and phase signal modulation by message x. [If a signal is modulated only with respect to amplitude, then $\psi(t, x) \equiv 0$; if it is modulated only with respect to phase, then $u(t, x) \equiv 1$]. Function u(t, x) is normalized so that

$$\overline{u^2(t,\ x)} = 1. \tag{18.36}$$

[In the case of a pulse signal, u(t, x) — impulse function with (where x = 0) spacing T_s. The assumption here is that signal pulses are coherent when there are no fluctuations ϕ(t)].

We will assume that signal fluctuations are relatively slow, i. e., function ρ (t) changes slowly compared with u(t, x) and ψ (t, x). However, even given such limitations placed on the nature of the signal fluctuations, resultant /338 calculations are more unwieldy than is the case for the aforementioned regular signal (see [127], for example). Therefore, we will present here only the most important result (obtained in [127]) — the formula for determination of the spectral density of the noise at discriminator output, given additive noise $u_ш(t)$ in the form of normal white noise:

$$G_{ш\,e} = \frac{2}{\gamma_e\left\{\overline{\left[\frac{\partial u(t,\ x)}{\partial x}\right]^2} + \overline{u^2(t,\ x)\left[\frac{\partial\psi(t,\ x)}{\partial x}\right]^2} - \left[\overline{u^2(t,\ x)\,\frac{\partial\psi(t,\ x)}{\partial x}}\right]^2\right\}}, \tag{18.37}$$

where

$$\gamma_e = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{S_\phi^2(\omega)/N_0^2}{1+\frac{1}{N_0}S_\phi(\omega)}\,d\omega; \tag{18.38}$$

$S_\phi(\omega)$ — signal fluctuation power spectrum (bilateral), linked with fluctuation correlation function ρ (t) by a Fourier transform:

$$\rho(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_c(\omega) \cos \omega\tau \, d\omega; \quad (18.39)$$

$$\rho(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_c(\omega) \, d\omega = P_{cp}. \quad (18.40)$$

If there is no phase modulation $\psi(t, x)$, then, the result from (18.37) is

$$g_{m0} = \frac{2}{\gamma_0 \overline{\left[\frac{\partial u(t, x)}{\partial x}\right]^2}}. \quad (18.41)$$

Formula (18.37) significance is that it makes it possible to determine, for a fluctuating signal, the magnitude of the spectral density of noise $\Delta \varepsilon_n(t)$ at optimum discriminator output, while this spectral density, as follows from Figure 18.3, determines potential message fidelity.

If signal fluctuation correlation function $\rho(\tau)$ has an exponential nature, the spectrum $S_c(\omega)$ has the form

$$S_c(\omega) = \frac{P_{cp}}{\Delta f_c} \frac{1}{1 + (\omega/2\Delta f_c)^2}, \quad (18.42)$$

where

$$\Delta f_c = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_0(\omega) \, d\omega \quad (18.43)$$

— equivalent width of fluctuation normalized spectrum $S_0(\omega)$, i. e., the spectrum determined from relationship /339

$$S_0(\omega) = \frac{S_c(\omega)}{S_c(0)} = \frac{1}{1 + \left(\frac{\omega}{2\Delta f_c}\right)^2}. \quad (18.44)$$

Substituting expression (18.42) into (18.36) provides

$$\gamma_0 = 2\Delta f_c h \left(1 - \frac{1}{\sqrt{1+h}}\right) = \frac{2\Delta f_c h^2}{(1 + \sqrt{1+h})\sqrt{1+h}}. \quad (18.45)$$

where

$$h = \frac{P_{cp}}{N_0 \Delta f_c} \tag{18.46}$$

is the ratio of signal mean power to the mean additive noise power in the band equalling equivalent width Δf_c of the signal fluctuation spectrum.

Given a high signal-to-noise ratio ($h \gg 1$), from (18.45) we obtain

$$\gamma_0 \approx 2\Delta f_c h = \frac{2P_{cp}}{N_0}. \tag{18.47}$$

Here, if there is no phase modulation $\psi(t, x)$, for g_{m0} formula (18.41) is valid, from which, considering (18.47), we obtain

$$g_{m0} = \frac{N_0}{P_{cp}\overline{\overline{\left[\frac{\partial u(t, x)}{\partial x}\right]^2}}}. \tag{18.48}$$

But, when there is no phase modulation $\psi(t, x)$, from (18.34) we obtain [considering E(t) slowness compared to u(t, x)]

$$\overline{\overline{\left[\frac{\partial u_c(t, x)}{\partial x}\right]^2}} = \overline{\overline{E^2(t)}}\ \overline{\overline{\left[\frac{\partial u(t, x)}{\partial x}\right]^2}}\ \frac{1}{2} = P_{cp}\overline{\overline{\left[\frac{\partial u(t, x)}{\partial x}\right]^2}}.$$

Therefore, formula (18.48) may be written also in the following form:

$$g_{m0} = \frac{N_0}{\overline{\overline{\left[\frac{\partial u_c(t, x)}{\partial x}\right]^2}}}. \tag{18.49}$$

This expression completely coincides with relationship (18.31) for a determinate signal.

Thus, for the aforementioned conditions [i. e., where $h \gg 1$, $\psi(t, x) \equiv 0$ and signal fluctuation slowness compared with u(t, x)], the spectral density g_{m0} value and, consequently, potential message x(t) fidelity as well, is obtained for a fluctuating signal, just as for a determinate signal with the same mean power. However, in the general case, signal fluctuations will lead, naturally, /340

to a deterioration in potential accuracy, i. e., to an increase in spectral density $g_{m\bullet}$ (given the same mean signal power).

18.3 Optimum Linear Filter Φ_1 (of Optimum Smoothing Networks) System Design and Determination of Potential Message Fidelity

As noted in § 18.1, optimum filter Φ_1 structure (Figure 18.1) completely is determined from relationships (18.4)-(18.6), while potential fidelity is determined from formula (18.13). A detailed study of these relationships for various message x(t) characteristics is presented in [127]. We will limit ourselves here for brevity to examination only of several most-characteristic cases.

We will begin with examination of a case when message x(t) may be considered a stationary random process and may be restricted to minimization of the root-mean-square error only for the steady-state mode, i. e, to assume that $t_0 = -\infty$. Here, optimum linear filter Φ_1 has constant parameters and, consequently, completely is characterized by its impulse response $\eta(t-\tau)$ or transfer function $K_1(j\omega)$. In this case, it is more convenient to find these characteristics by reducing the problem to optimum linear filtration, examined in § 2.2, rather than direct solution of integral equations (18.4) and (18.5).

It is evident from the Figure 18.3b and 18.3c circuits that equivalent filters Φ_2 and $\Phi_\bullet$ have the following transfer functions, respectively

$$K_2(j\omega) = kK_1(j\omega) \tag{18.50}$$

and

$$K_\bullet(j\omega) = \frac{K_2(j\omega)}{1+K_2(j\omega)}. \tag{18.51}$$

The Figure 18.3c circuit completely coincides with the optimum linear filtration circuit Wiener examined and which was presented in § 2.2. Therefore, formula (2.14) is valid for optimum transfer function $K_\bullet(j\omega)$ system design. Since simple message reproduction [h(t) = x(t)] is required in this case, while message x(t) and noise $\Delta e_n(t)$ may be considered statistically independent, then

$$S_{yh}(\omega) = S_x(\omega); \quad S_y(\omega) = S_x(\omega) + S_m(\omega), \tag{18.52}$$

and formula (2.14) provides

$$K_0(j\omega)=\frac{1}{2\pi\psi(j\omega)}\int_0^{\infty}e^{-j\omega t}\,dt\int_{-\infty}^{\infty}\frac{S_{x}(\omega)}{\psi(-j\omega)}e^{j\omega t}\,d\omega, \tag{18.53}$$

where

$$\psi(j\omega)\,\psi(-j\omega)=|\psi(j\omega)|^{2}=S_{y}(\omega). \tag{18.54}$$

Power spectrum $S_y(\omega)$ may, without a significant loss of generality, /341
be represented in the form of the rational function

$$S_y(\omega)=\frac{|Q_m(j\omega)|^m}{|P_n(j\omega)|^n}, \tag{18.55}$$

where $Q_m(j\omega)$ and $P_n(j\omega)$ -- polynomials of degrees m and n (where n $\geqslant$ m) of $j\omega$ with weight factors. Here, functions $\psi(j\omega)$ and $\psi(-j\omega)$ included in expression (18.53) may be found by the method of special expansion (factorization) of function $S_y(\omega)$ into factors with zeros and pluses in the upper half-plane [at $\psi(j\omega)$] and in the lower half-plane [at $\psi(-j\omega)$] of complex variable ω (see [127], for instance).

For practical use, it usually is more convenient to represent formula (18.53) in the following form [127]:

$$K_0(j\omega)=\frac{1}{\psi(j\omega)}\left[\frac{S_x(\omega)}{\psi(-j\omega)}\right]_+, \tag{18.56}$$

where the operation $[\,]_+$ denotes the taking of that term of the expression in brackets which has pluses in the upper half-plane of complex variable ω.

It follows from formulas (18.50) and (18.51) that

$$K_1(j\omega)=\frac{1}{k}\,\frac{K_0(j\omega)}{1-K_0(j\omega)}. \tag{18.57}$$

Therefore, having computed magnitude $K_0(j\omega)$ from formula (18.56), it then is possible to find needed filter Φ_1 optimum transfer function $K_1(j\omega)$ from formula (18.57).

The methodology Wiener demonstrated for finding optimum transfer function $K_0(j\omega)$ insures physical realizability of this transfer function. Filter $K_1(j\omega)$ turns out also to be physically realizable when frequency spectrum $S_y(\omega)$ is represented in a (18.55)-type of rational function and, consequently, the methodology presented above for its system design is valid.

As our example, we will examine a case when the message x(t) spectrum may be represented in the form

$$S_x(\omega)=\frac{2\sigma_0^2 T}{1+(\omega T)^2}, \tag{18.58}$$

where σ_0^2 — variance of magnitude x, while 1/T — width of the spectrum at the half-power level.

Since, in the case being examined, $\Delta e_n(t)$ — white noise with unilateral spectral density g_{m0} = 2k, then its bilateral spectral density equals

$$S_{m0}=g_{m0}/2=1/k, \tag{18.59}$$

and, considering (18.52) and (18.59), we have

$$S_y(\omega)=\frac{g_{m0}[1+\rho+(\omega T)^2]}{2[1+(\omega T)^2]}, \tag{18.60}$$

where /342

$$\rho=\frac{4\sigma_0^2 T}{g_{m0}}. \tag{18.61}$$

Next, in accordance with (18.54), we will expand function $S_y(\omega)$ into factors $\psi(j\omega)$ and $\psi(-j\omega)$ containing zeros and pluses, respectively, in the upper and in the lower half-planes of frequency ω, considered a complex variable. In the given case, this expansion is evident:

$$S_y(\omega)=\frac{g_{m0}}{2}\,\frac{(\sqrt{1+\rho}+j\omega T)(\sqrt{1+\rho}-j\omega T)}{(1+j\omega T)(1-j\omega T)}. \tag{18.62}$$

Therefore, it is possible to assume

$$\psi(j\omega)=\frac{g_{m\,e}}{2}\frac{(\sqrt{1+\rho}+j\omega T)}{(1+j\omega T)};\quad \psi(-j\omega)=\frac{\sqrt{1+\rho}-j\omega T}{1-j\omega T}. \tag{18.63}$$

In accordance with (18.56), we will find

$$\left[\frac{S_x(\omega)}{\psi(-j\omega)}\right]_+=\frac{g_{m\,e}\rho}{2}\left[\frac{1-j\omega T}{[1+(\omega T)^2](\sqrt{1+\rho}-j\omega T)}\right]_+=$$
$$=\left[\frac{\rho g_{m\,e}}{2(1+j\omega T)(\sqrt{1+\rho}-j\omega T)}\right]_+=\frac{\rho g_{m\,e}}{2(\sqrt{1+\rho}+1)}\times$$
$$\times\left[\frac{1}{1+j\omega T}+\frac{1}{\sqrt{1+\rho}-j\omega T}\right]_+=\frac{\rho g_{m\,e}}{2(\sqrt{1+\rho}+1)(1+j\omega T)}. \tag{18.64}$$

Substituting relationships (18.63) and (18.64) into (18.56), we obtain

$$K_0(j\omega)=\frac{\rho}{(\sqrt{1+\rho}+1)(\sqrt{1+\rho}+j\omega T)}. \tag{18.65}$$

Here, formula (18.57) provides

$$K_1(j\omega)=\frac{g_{m\,e}\rho}{2(\sqrt{1+\rho}+1)}\frac{1}{(1+j\omega T)}. \tag{18.66}$$

Consequently, in the case under examination, optimum filter Φ_1 must comprise an amplifier with gain

$$K_0=\frac{\rho g_{m\,e}}{2(\sqrt{1+\rho}+1)}=\frac{2\sigma_0^2 T}{\left(\sqrt{1+\frac{4\sigma_0^2 T}{g_{m\,e}}}+1\right)} \tag{18.67}$$

and an inertia component with constant time T.

We now will find potential message measurement accuracy determined from relationship (18.13):

$$\sigma_\varepsilon^2=\overline{\varepsilon^2(t)}=C(t,t). \tag{18.13}$$

As noted in § 18.1, $kC(t,\tau)$ is the filter Φ_0 impulse response (Figure 18.3c). In the case being examined, this filter has constant parameters and, consequently, $C(t,\tau)=C(t-\tau)$. Therefore, $\sigma_\varepsilon^2 = C(0)$.

Considering that a Fourier transform links the filter Φ_φ impulse response $kC(t - \mathcal{T})$ with its transfer function $K_\varphi(j\omega)$, the following relationship may be obtained [127]:

$$kC(0) = \lim_{\omega\to\infty} j\omega K_\varphi(j\omega). \tag{18.68}$$

Therefore, /343

$$\sigma_\varepsilon^2 = \frac{1}{k}\lim_{\omega\to\infty} j\omega K_\varphi(j\omega), \tag{18.69}$$

and, considering (18.61) and (18.65), we have

$$\left.\begin{aligned} &\sigma_\varepsilon^2 = \frac{\rho}{kT(\sqrt{1+\rho}+1)} = \frac{2\sigma_\varphi^2}{\sqrt{1+\rho}+1}, \\ \text{where}\quad &\rho = \frac{4\sigma_\varphi^2 T}{g_{m\varphi}}. \end{aligned}\right\} \tag{18.70}$$

For comparison, we now will examine solution of the same problem using a simplified Bode-Shannon method, in which no condition of physical accomplishment is placed on the filter system being designed. Here, in accordance with formulas (2.19) and (2.20), we have

$$K_\varphi(j\omega) = \frac{1}{1 + \dfrac{S_y(\omega)}{S_x(\omega)}} \tag{18.71}$$

and

$$\sigma_\varepsilon^2 = \overline{\varepsilon^2_{\min}} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S_m(\omega)\,d\omega}{1 + \dfrac{S_m(\omega)}{S_x(\omega)}}. \tag{18.72}$$

Substitution of expressions (18.58) and (18.60) into these formulas provides

$$K_\varphi(j\omega) = \frac{\rho}{1+\rho+(\omega T)^2} \tag{18.73}$$

and

$$\sigma_\varepsilon^2 = \frac{\sigma_\varphi^2}{\sqrt{1+\rho}}. \tag{18.74}$$

It follows from (18.57), (18.59), and (18.73) that

$$K_1(j\omega) = \frac{\rho g_{m0}}{2}\,\frac{1}{1+(\omega T)^2}. \tag{18.75}$$

As is evident, resultant transfer function $K_1(j\omega)$ is not physically realizable, but the potential accuracy value found here (18.70) slightly differs from corresponding value (18.70) obtained for a physically-realizable filter. This example confirms the validity of the assertion made in § 2.2 that the simplified Bode-Shannon method is applicable for approximate estimation of potential message fidelity.

The aforementioned analysis was performed for a case when reproduced message x(t) may be considered a stationary random process. However, in a number of cases, message x(t) reproduction itself, but only its first or higher derivative with respect to time, may be considered a stationary process. For example, if aircraft acceleration changes randomly and stationarily, then its velocity (an integral of acceleration) is a random process with a stationary first derivative, while its coordinate is a random process with a stationary second derivative. Processes of this type fall in the class of so-called processes with stationary increments (stationary derivatives).

The methodology for design of linear filter systems for such processes /344 is presented in [127]. We will limit ourselves to examination only of the simplest case when message x(t) may be approximated by the integral of stationary white noise (such a process also often is called a Wiener process).

The correlation function of such a process equals

$$R_x(t,\tau) = \left\{ \begin{array}{ll} Bt & \text{where } t < \tau; \\ B\tau & \text{where } t > \tau. \end{array} \right\} \tag{18.76}$$

The resultant solution of equations (18.4) and (18.5) for this case in [127] provides the following results:

$$C(t,\tau)=\sqrt{\frac{B}{k}}\,\frac{\operatorname{sh}\sqrt{Bk}\tau}{\operatorname{ch}\sqrt{Bk}t}; \tag{18.77}$$

$$\eta(t,\tau)=\sqrt{\frac{B}{k}}\,\operatorname{th}\tau. \tag{18.78}$$

Substituting these relationships into expressions (18.3) and (18.13), we obtain

$$\gamma(t)=\int_{t_0}^{t}\sqrt{\frac{B}{k}}\,(\operatorname{th}\tau)\,z(\tau)\,d\tau+\overline{x(t)} \tag{18.79}$$

and

$$\sigma_s^2=\sqrt{\frac{B}{k}}\,\operatorname{th}(\sqrt{Bk}t), \tag{18.80}$$

where $k \doteq 2/g_{so}$.

It follows from (18.78) and (18.79) that, in the case being examined, the linear filter must comprise discriminator z(t) output voltage amplifier, with amplification changing over time in accordance with the law of the hyperbolic tangent, i. e., with a rise from zero to $\sqrt{B/k}$), and a subsequent integrator. Here, as follows from (18.80), the message reproduction error variance also rises in accordance with the law of the hyperbolic tangent, striving towards steady-state value $\sqrt{B/k}$.

As our final example, we will examine a case when message x(t) may be represented in the form of a linear combination of known functions with random factors, i. e., in the form

$$x(t)=\overline{x(t)}+\sum_{k=1}^{m}A_k f_k(t). \tag{18.81}$$

For the simplest case, when

$$x(t)=\overline{x(t)}+Af(t) \tag{18.82}$$

and parameter A has a normal law of distribution with a zero mean value and variance

$$\sigma_0^2=\overline{A^2}, \tag{18.83}$$

the following results are obtained in [127]:

$$C(t,\tau)=\frac{\sigma_0^2 f(t) f(\tau)}{1+k\sigma_0^2\int_{t_0}^{t} f^2(s)\,ds} \tag{18.84}$$

and /345

$$\eta(t,\tau)=\frac{\sigma_0^2 f(t) f(\tau)}{1+k\sigma_0^2\int_{t_0}^{\tau} f^2(s)\,ds}. \tag{18.85}$$

Here, in accordance with formulas (18.3) and (18.13), we have

$$\gamma(t)=\int_{t_0}^{t}\frac{\sigma_0^2 f(t) f(\tau) z(\tau)}{1+k\sigma_0^2\int_{t_0}^{\tau} f^2(s)\,ds}\,d\tau+\overline{x(t)}. \tag{18.86}$$

$$\sigma_\varepsilon^2=\frac{\sigma_0^2 f^2(t)}{1+\frac{2\sigma_0^2}{g_m e}\int_{t_0}^{t} f^2(s)\,ds}. \tag{18.87}$$

The optimum filter Φ_1 structural schematic depicted in Figure 18.8, it is easy to see, corresponds to expression (18.86). This circuit will comprise two multipliers, adder, integrator, and three generators of known time functions $u_1(t)$, $u_2(t)$, and $u_3(t)$.

It follows from formula (18.87) that measurement error variance and the error itself will become equal to zero in the steady-state mode (where $t \longrightarrow \infty$). This is because, in the case under consideration, the only unknown is message x(t) constant parameter A and, when $t \longrightarrow \infty$, it may be measured precisely.

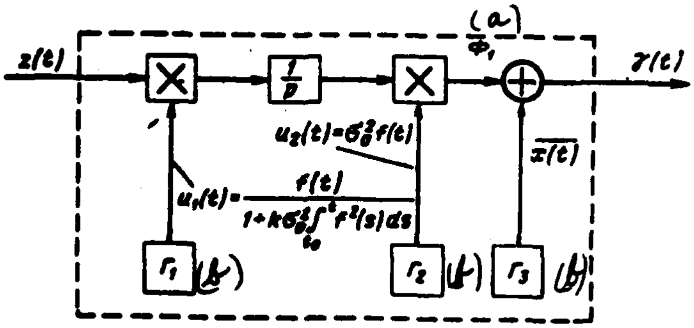


Figure 18.8. (a) -- F_1 [Filter]
(b) -- G_1--G_3 [Generators].

The examples presented above for design of a linear filter system (for smoothing networks) and for determination of potential message fidelity clearly confirm the following general maxims:

1. The structure of an optimum linear filter and its parameters very significantly will depend on the nature of reproduced message x(t) and on whether high message fidelity is required only in the steady-state mode (where $t_0 = -\infty$ or $t \to \infty$) or is required in the transient mode as well.

2. Optimum linear filter parameters will depend on the magnitude of the spectral density $g_{ш.э}$ of the noise at discriminator output, which, in turn, /346 will depend on the signal power to noise power ratio. Consequently, filter parameters must change when this ratio changes. No such change is envisioned in the optimum system examined above (Figure 18.1) since, during its system design, all a priori signal and noise characteristics were assumed known. Therefore, in an actual system operating under conditions whereby many of these a priori characteristics are unknown, the accuracy in the general case will be lower than its potential accuracy found above, even if you envisage in this system devices to measure noise and signal parameters (because such measurements may not be precise due to the noise action).

3. Potential message fidelity, characterized by the magnitude of error variance σ_e^2 , monotonically will depend on the spectral density $g_{ш.э}$ of the noise at

discriminator output. The variance will strive monotonically towards zero when spectral density $g_{m\phi}$ strives towards zero.

Such a result is understandable since, in the optimum system structural schematic (Figure 18.2), noise $\Delta e_n(t)$ with spectral density $g_{m\phi}$ is the single primary source of message reproduction errors; when this noise is absent ($g_{m\phi} = 0$), message x(t) reproduction may be as precise as desired, given a sufficiently-broad filter bandwidth.

Hence, it follows that spectral density $g_{m\phi}$, characterizing the potential accuracy of discriminator action, is a vital parameter determining the potential accuracy of system operation as a whole.

18.4 Potential Accuracy Given Complex Message Reproduction and Indirect Modulation Types

The results presented in § 18.3 applied to direct signal modulation types and to simple message reproduction. We now will explain the special features which arise for indirect modulation types and during complex message reproduction. Here, we will restrict ourselves to examination of the problem of potential determinate signal fidelity.

It was demonstrated in § 18.2 that the presence of signal fluctuations during the same signal mean power may cause deterioration in message fidelity or keep it essentially unchanged. Therefore, the potential accuracy value found for a determinate signal may be considered threshold potential accuracy, which may not be exceeded regardless of actual signal shape.

During indirect modulation, oscillation y(t) at measurement system input has the following form:

$$y(t) = u_c(t, x_1) + u_m(t), \tag{18.88}$$

where

$$x_1(t) = \mathcal{D}_1(p)\, x(t). \tag{18.89}$$

Here, $\mathcal{D}_1$ (p) -- linear differential operator; p = d/dt -- elementary differen- /347
tiation operator; x(t) -- measured parameter; $x_1(t)$ -- direct parameter, i. e., parameter included in function $u_c(t, x_1)$ directly, rather than under the sign of the operator.

Measurement of signal frequency Doppler shift $F_Д(t)$ is an example of indirect modulation. In this case

$$y(t)=U_0\cos[\omega_0 t+2\pi\int F_Д(t)\,dt+\varphi_0]+u_m(t). \qquad (18.90)$$

Here, it is evident that

$$\left.\begin{aligned} x(t) &= F_Д(t),\\ x_1(t) &= \int F_Д(t)\,dt = \frac{1}{p}x(t),\\ \mathcal{D}_1(p) &= 1/p. \end{aligned}\right\} \qquad (18.91)$$

The requirement in simple measurement is to measure parameter x(t) with minimum root-mean-square error, i. e., to insure that this condition is met

$$\overline{\varepsilon^2(t)}=\overline{[x(t)-\gamma(t)]^2}=\min, \qquad (18.92)$$

where $\gamma(t)$ -- measurement result.

The requirement in complex measurement is to measure with minimum root-mean-square error, not parameter x(t) itself, but some function $x_0(t)$, linked with x(t) by an operator relationship. For simplicity, we will consider the operator linear, i. e., we will assume

$$x_0(t)=\mathcal{D}_0(p)\,x(t). \qquad (18.93)$$

So, during complex measurement, the requirement is to meet this condition

$$\overline{\varepsilon_0^2(t)}=\overline{[x_0(t)-\gamma(t)]^2}=\min, \qquad (18.94)$$

where $\gamma(t)$ -- measurement result.

Consequently, given indirect modulation types and complex measurement, there are three types of parameters--x(t), $x_1(t)$, and $x(t)$, $x_1(t)$. In future, we will call x(t) the primary parameter, $x_1(t)$ the direct parameter, and $x_э(t)$ the equivalent parameter.

It follows from relationships (18.89) and (18.93) that

$$x_э(t) = \mathfrak{D}(p)\, x_1(t), \tag{18.95}$$

where

$$\mathfrak{D}(p) = \mathfrak{D}_1^{-1}(p)\, \mathfrak{D}_2(p), \tag{18.96}$$

$$\mathfrak{D}_1^{-1}(p) = 1/\mathfrak{D}_1(p) \tag{18.97}$$

-- linear operator inverse with respect to $\mathfrak{D}_1$ (p).

It follows from expressions (18.88), (18.94), and (18.95) that, during indirect modulation and complex measurement, the requirement is to measure with minimum root-mean-square error, not direct parameter $x_1(t)$ included directly in the signal function, but some equivalent parameter $x_э(t)$, linked with $x_1(t)$ by linear operator $\mathfrak{D}$(p).

Expression (18.95) corresponds to a general case when both indirect /348 modulation and complex measurement occur simultaneously. It encompasses the following particular cases:

1. Indirect modulation, but simple measurement. Here, evidently,

$$\mathfrak{D}_2(p) = 1 \quad \text{and} \quad \mathfrak{D}(p) = \mathfrak{D}_1^{-1}(p). \tag{18.98}$$

2. Direct modulation, but complex measurement. Here

$$\mathfrak{D}_1(p) = 1 \quad \text{and} \quad \mathfrak{D}(p) = \mathfrak{D}_2(p). \tag{18.99}$$

3. Direct modulation and simple measurement. In this case

$$\mathcal{D}_1(p) = \mathcal{D}_2(p) = \mathcal{D}(p) = 1. \tag{18.100}$$

It follows from what has been stated that the general case (indirect modulation and complex measurement) differs from the particular case (direct modulation and simple measurement) examined in preceding sections only in that equivalent parameter $x_э(t)$, linked with direct parameter $x_1(t)$ by linear operator $\mathcal{D}(p)$, rather than the direct parameter, will be subject to measurement with minimum root-mean-square error.

We will clarify which changes must be introduced here into the equivalent system circuit that determines potential measurement accuracy.

This circuit may be represented in the form depicted in Figure 18.9 for direct modulation and simple measurement examined in preceding sections. The task of

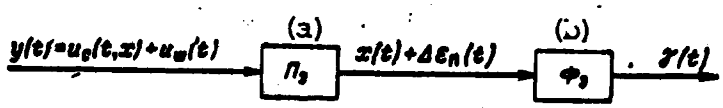


Figure 18.9. (a) -- Equivalent receiver;
(b) -- Equivalent filter.

equivalent optimum receiver $П_э$ is reproduction of direct parameter x(t) with the guarantee of minimum noise $\Delta\varepsilon_п(t)$ spectral density $g_{ш э}$. Here, receiver $П_э$ structure, noise $\Delta\varepsilon_п(t)$ character, and magnitude $g_{ш э}$ will not depend on reproduced parameter x(t) type; only the structure of equivalent optimum linear filter $Φ_э$ and the magnitude of the root-mean-square error will depend on this parameter (these results are valid if the magnitude of the root-mean-square error is sufficiently small).

Hence, it follows that, in the general case, when the requirement is measurement of equivalent parameter $x_э(t)$, rather than direct parameter $x_1(t)$ itself, with minimum root-mean-square error, potential measurement accuracy may be determined with the aid of the equivalent circuit depicted in Figure 18.10. In this /349
circuit, receiver $П_э$ and noise $\Delta\varepsilon_п(t)$ are identical to those in Figure 18.9.

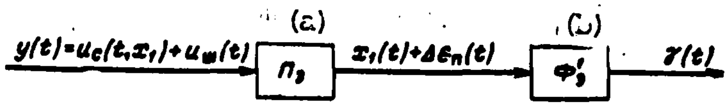


Figure 18.10. (a) -- Equivalent receiver;
(b) -- Equivalent filter.

Actually, let primary parameter x(t) be identical in both circuits. Then, in the Figure 18.9 circuit, receiver $\Pi_э$ extracts parameter x(t), while in the Figure 18.10 circuit, receiver $\Pi_э$ extracts parameter $x_1(t)$ linked with x(t) by relationship (18.89) and, consequently, having a different character. But, as noted above, receiver $\Pi_э$ structure and noise $\Delta\varepsilon_п(t)$ at its output will not depend on the nature of the measured parameter (if this parameter's measurement error is sufficiently small). Therefore, it is possible to assert that receiver $\Pi_э$ and noise $\Delta\varepsilon_п(t)$ in Figure 18.10 may have exactly the same form as in Figure 18.9 (if, in accordance with problem conditions, it is possible to assume that the reproduction error is sufficiently small).

In this case, filter $\Phi_э'$ structure must differ since, in the Figure 18.9 circuit, when noise $\Delta\varepsilon_п(t)$ is present, the requirement is to insure minimum root-mean-square deviation of output oscillation $\gamma(t)$ from input oscillation x(t). In the Figure 18.10 circuit, the requirement is, given the same noise $\Delta\varepsilon_п(t)$, to insure minimum root-mean-square deviation of output oscillation $\gamma(t)$, not from input oscillation $x_1(t)$, but from oscillation $x_э(t)$ linked with $x_1(t)$ by relationship (18.95).

So, in the general case under discussion, it is possible during determination of potential measurement accuracy to assume that noise $\Delta\varepsilon_п(t)$ at linear network input has the form of normal white noise with spectral density $g_{mэ}$ determined from formula (18.31), in which x is replaced by x_1, i. e.,

$$g_{mэ} = \frac{N_0}{\overline{\left[\frac{\partial u_c(t,\, x_1)}{\partial x_1}\right]^2}}. \tag{18.101}$$

Here, optimal linear filter $\Phi_э'$ structure will be found from the optimum message $x_1(t)$ complex reproduction condition, i. e., from the condition

$$\overline{\varepsilon_э^2(t)} = \overline{[\gamma(t) - \mathcal{D}(p)\, x_1(t)]^2} = \min. \tag{18.102}$$

Having determined filter $\Phi_э'$ structure, it then is possible with the aid of the Figure 18.10 circuit to compute the magnitude of the measurement mean-square error, $\overline{\varepsilon_э^2(t)}$.

If you consider that, in accordance with (18.95)

$$x_э(t) = \mathcal{D}(p)\, x_1(t), \tag{18.103}$$

then it is possible to convert the Figure 18.10 equivalent circuit to the form depicted in Figure 18.11. In this circuit

$$\Delta\varepsilon_n'(t) = \mathcal{D}(p)\, \Delta\varepsilon_n(t), \tag{18.104}$$

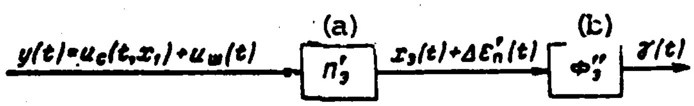


Figure 18.11. (a) -- Equivalent receiver;
(b) -- Equivalent filter;

while linear filter $\Phi_э''$ must insure that this condition is met /296

$$\overline{\varepsilon_э^2(t)} = \overline{[\gamma(t) - x_э(t)]^2} = \min, \tag{18.105}$$

i. e., provide optimum simple reproduction of message $x_э(t)$ arriving at its input, given additive independent noise $\Delta\varepsilon_n'(t)$ having the form of normal non-white noise with power spectrum

$$g'_{ш\,0}(f) = |\mathcal{D}(j2\pi f)|^2\, g_{ш\,0}. \tag{18.106}$$

In some cases, the Figure 18.10 circuit may turn out to be more convenient for determination of potential accuracy, while the Figure 18.11 circuit is more convenient in others.

For example, let the requirement be to find the potential accuracy of simple

measurement of frequency Doppler shift $F_Д(t)$. In this case, $\mathcal{D}_2(p) = 1$, while $\mathcal{D}_1(p) = 1/p$. Therefore, from (18.96) we have $\mathcal{D}(p) = p$. Then, the result from (18.101) and (18.106) is

$$g'_{m0}(f) \doteq \frac{N_0 (2\pi f)^2}{\left[\frac{\partial u_0(t, x_1)}{\partial x_1}\right]^2}, \tag{18.107}$$

where $x(t) = F_Д(t)$.

For illustration, we now will examine computation of potential measurement accuracy in a case when primary message x(t) is a normal stationary process with known mean value x(t) and power spectrum $\partial_x(f)$, and no condition of physical realizability is levied on the linear networks.

In this case, filters Φ_0' and Φ_0'' have constant parameters and are completely characterized by their transfer functions $K_0'(j\omega)$ and $K_0''(j\omega)$, respectively.

Here, we will begin from the Figure 18.11 circuit. Then, the problem boils down to optimum simple reproduction of message $x_0(t)$ on a background of normal noise $\Delta e_n'(t)$, where

$$\left.\begin{aligned} x_0(t) &= \mathcal{D}_2(p)\, x(t), \\ \Delta e_n'(t) &= \mathcal{D}(p)\, \Delta e_n(t) = \mathcal{D}_1^{-1}(p)\, \mathcal{D}_2(p)\, \Delta e_n(t). \end{aligned}\right\} \tag{18.108}$$

We will assume that operators $\mathcal{D}_1(p)$ and $\mathcal{D}_2(p)$ are such that processes $x_0(t)$ and $\Delta e_n'(t)$ may be considered stationary. Then, potential accuracy in the steady-state mode (where $t_0 = -\infty$) is determined from relationship (2.20) and

$$\overline{\varepsilon_0^2(t)} = \int_0^\infty \frac{g'_{m0}(f)\, df}{1 + \frac{g'_{m0}(f)}{g_{x_0}(f)}}, \tag{18.109}$$

where $g_{x_0}(f)$ and $g'_{m0}(f)$ -- unilateral power spectra of processes $x_0(t)$ and $\Delta e'_n(t)$, respectively.

It follows from (18.108) that

$$\left.\begin{array}{c} g_{x_0}(f) = |\mathcal{D}_2(j2\pi f)|^2 g_x(f), \\ g'_{\text{ш}\,0}(f) = |\mathcal{D}(j2\pi f)|^2 g_{\text{ш}\,0} = |\mathcal{D}_1^{-1}(j2\pi f)|^2 |\mathcal{D}_2(j2\pi f)|^2 g_{\text{ш}\,0}. \end{array}\right\} \tag{18.110}$$

Therefore /351

$$\overline{\varepsilon_0^2} = \int_0^\infty \frac{|\mathcal{D}(j2\pi f)|^2 g_{\text{ш}\,0}\,df}{1 + \dfrac{g_{\text{ш}\,0}}{|\mathcal{D}_1(j2\pi f)|^2 g_x(f)}}. \tag{18.111}$$

As our example, we will examine measurement of signal frequency Doppler shift $F_Д(t)$. Here, as noted above,

$$\mathcal{D}_1(p) = 1/p, \quad \mathcal{D}_2(p) = 1, \quad \mathcal{D}(p) = p, \tag{18.112}$$

and expressions (18.108) take the form

$$x_0(t) = x(t), \tag{18.113}$$

$$\Delta e_{\text{ш}}'(t) = p\Delta e_{\text{ш}}(t). \tag{18.114}$$

Since we assumed that x(t) is a normal stationary process, then $x_0(t)$ is the same type process. Noise $\Delta e_{\text{ш}}(t)$ is normal stationary white noise. Strictly speaking, there is no derivative of such a process. Therefore, in order to eliminate this difficulty during the operation with noise $\Delta e_{\text{ш}}'(t)$, it should be assumed that white noise $\Delta e_{\text{ш}}(t)$ has a spectrum bound by some high frequency $f_в$. Since selected frequency $f_в$ may be as high as desired (but finite), then this will not impact upon computational results if spectrum $g_x(f)$ in the region of high frequencies decreases rapidly enough. Here, the derivative, $\Delta e_{\text{ш}}'(t)$, also is a normal stationary process and, consequently, assumptions made during derivation of formula (18.111) are satisfied.

Substituting relationship (18.112) into formula (18.111), we obtain

$$\overline{\varepsilon_0^2} = \int_0^\infty \frac{4\pi^2 g_{\text{ш}\,0} f^2\,df}{1 + \dfrac{4\pi^2 g_{\text{ш}\,0} f^2}{g_x(f)}}. \tag{18.115}$$

where g_{m0} is determined from formula (18.101). Since, in this case

$$u_0(t, x_1) = U_0 \cos[\omega_0 t + 2\pi x_1(t) + \varphi_0],$$

then

$$B = 4\pi^2 \frac{U_0^2}{2} = 4\pi^2 P_{cp},$$

where $P_{cp} = U_0^2/2$ -- specific signal mean power. Therefore, formula (18.101) provides

$$g_{m0} = \frac{N_0}{4\pi^2 P_{cp}}. \qquad (18.116)$$

Relative measurement error may be determined from the formula

$$\frac{\overline{\varepsilon_0^2}}{\sigma_0^2} = \frac{1}{\sigma_0^2}\int_0^\infty \frac{4\pi^2 g_{m0} f^2\, df}{1 + \frac{4\pi^2 g_{m0} f^2}{g_x(f)}}. \qquad (18.117)$$

Here, σ_0^2 -- measured parameter x(t) variance, i. e.,

$$\sigma_0^2 = \int_0^\infty g_x(f)\, df, \qquad (18.118)$$

where $x(t) = F_Д(t)$.

It is easy to become convinced that formula (18.115) coincides (precise /352
to designations) with formula (7.34) if the latter is used for an FM signal. Actually, for an FM signal, $E_m^2(f)$ is determined from formula (7.15), in which one should assume $\psi_x(t) = x_1(t)$. Then, considering (18.101), the result is

$$E_m^2(f) = \frac{N_0 (2\pi f)^2}{\left[\frac{\partial u_0(t, x_1)}{\partial x_1}\right]^2} = g_{m0}(2\pi f)^2. \qquad (18.119)$$

If, in addition, you consider that, in Chapter 7, message x(t) was assumed to be normalized so that $x_{макс} = 1$, while this normalization is missing in this case,

then message spectra $E_x^2(f)$ and $g_x(f)$ must coincide when $U_{\text{м}\Omega} = 1$. Considering this comment and relationship (18.119), the result is that formulas (18.115) and (7.34) actually coincide completely.

However, the formulas obtained in Chapter 7 based on the Kotel'nikov theory applied to a case of simple message reproduction and only one of the direct modulation types--frequency modulation. The material presented in this section makes it possible to determine potential accuracy for various types of indirect modulation and during complex message reproduction.

18.5 Multichannel Reception

The assumption in all preceding chapters was that single signal-plus-noise realization y(t) arrives at receiving device input and this realization will comprise information only about single usable message x(t). However, in practical cases, more-complex instances often will be encountered in which 1 different messages $x^{(1)}(t), \ldots, x^{(l)}(t)$ will be subject to reproduction, i. e., the receiving device must have an l-channel, rather than single-channel, output. In addition, this combination of messages [or single message x(t)] may comprise, not one input realization y(t), but a combination $\{y^{(1)}(t), \ldots, y^{(m)}(t)\}$ of such realizations, which arrive from m different antennas (or from m primary message processing devices, if we are talking about system design of secondary-processing devices), for example. In this case, the receiving device has m input channels. Consequently, in the general case, the optimum receiving system being designed may comprise m inputs and 1 outputs. The theory of optimum nonlinear filtration presented above also permits generalization for this more-complicated case.

From the point of view of the principle (idea) involved, such generalization is not complicated. Actually, the combination of input realizations may be considered some vector realization

$$\overrightarrow{y(t)} = \{y^{(1)}(t), \ldots, y^{(m)}(t)\}, \tag{18.120}$$

while the combination of reproduced messages--as some vector message

$$\overrightarrow{x(t)} = \{x^{(1)}(t), \ldots, x^{(l)}(t)\}. \tag{18.121}$$

It is possible to consider the result of reproduction of these messages some /353 vector

$$\overrightarrow{\gamma(t)} = \{\gamma^{(1)}(t), \ldots, \gamma^{(l)}(t)\}. \tag{18.122}$$

It is possible to introduce the concept of generalized loss function $l(\vec{x}, \vec{\gamma})$, which designates losses corresponding to each combination of vectors $\vec{x}$ and $\vec{\gamma}$. Here, it is possible to demonstrate [127] that, in the case of a generalized quadratic loss function analog, as usual, relationship (17.46) determines optimum receiver structure; the only requirement is to replace x, y, and γ in it with $\vec{x}$, $\vec{y}$, and $\vec{\gamma}$, respectively. However, reduction of expression (17.46) to a form applicable for practical use, as noted above, turns out to be complicated, even for the single-channel case (m = 1, l = 1). It is natural that this complexity rises sharply during conversion to a general case of m input and l output channels. Several useful results for this general case were obtained in [127].

In this book, we will restrict ourselves to examination of a very common, but also very important, case when l = 1, but the m signal-plus-noise input realizations have the following form:

$$\left.\begin{array}{l} y^{(1)}(t) = u_{c1}(t, x) + u_{m1}(t), \\ y^{(2)}(t) = u_{c2}(t, x) + u_{m2}(t), \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ y^{(m)}(t) = u_{cm}(t, x) + u_{mm}(t), \end{array}\right\} \tag{18.123}$$

where $u_{ci}(t, x)(i = 1, \ldots, m)$ -- precisely-known signals (except for message x(t), while $u_{mi}(t)(i = 1, \ldots, m)$ -- independent normal white noise with spectral densities $N_{01}, N_{02}, \ldots, N_{0m}$, respectively.

This case differs from those in preceding sections of the book only in that there is a requirement to examine an (18.120)-type vector realization, rather than single realization y(t). Therefore, formally, all results obtained above will remain valid if $\overrightarrow{y(t)}$ replaces y(t) in them.

In particular, the optimum message x(t) reproduction system structure depicted in Figure 18.1 takes the form shown in Figure 18.12. The only difference is that

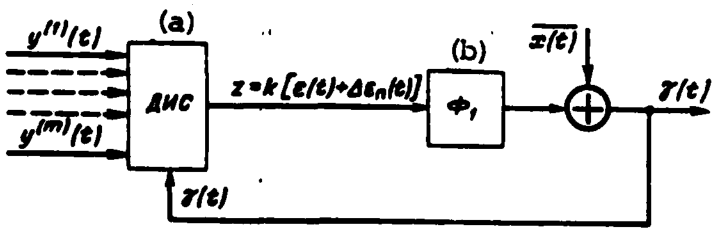


Figure 18.12. (a) -- DIS [Discriminator];
(b) -- F_1 [Filter].

a combination $\{y^{(1)}(t), \ldots, y^{(m)}(t)\}$ of realizations, rather than one realization y(t), arrives at discriminator input. All formulas determining this system's structure remain the same (as in a case where m = 1 as well), with the exception that, in expressions (18.6)÷(18.8), $\vec{y}$ should replace y, i. e., one should assume

$$z(t) = -\left[\frac{\partial Q(\vec{y}, x, t)}{\partial x}\right]_{x=\gamma}, \tag{18.124}$$

/354

$$k = \overline{\left[\frac{\partial^2 Q(\vec{y}, x, t)}{\partial x^2}\right]}, \tag{18.125}$$

where function $Q(\vec{y}, x, t)$ is determined from the relationship

$$\int_{t-\Delta t}^{t} Q(\vec{y}, x, t)\, dt = -\ln P_x(\vec{y}). \tag{18.126}$$

Hence, it follows that it suffices to compute the value of likelihood function $P_x(\vec{y})$ in order to consider the special features of a multichannel case ($m > 1$) compared to a single-channel case ($m = 1$).

Let the laws of noise distribution (multidimensional probability densities) $u_{m1}(t), \ldots, u_{mm}(t)$ included in expression (18.123) equal $W_1(u_{m1}), \ldots, W_m(u_{mm})$, respectively.

Then, considering that this noise is interdependent, while signals $u_{c1}(t, x)$, . . ., $u_{cm}(t, x)$ for a given x are precisely known, we will obtain

$$P_x(\vec{y}) = P_x(y^{(1)}, \dots, y^{(m)}) = \\ = W_1(y^{(1)} - u_{c1}) \dots W_m(y^{(m)} - u_{cm}). \quad (18.27)$$

But, law of distribution $W_i(u_{\text{ш}i})$ of normal white noise $u_{\text{ш}i}(t)$ has the form

$$W_i(u_{\text{ш}i}) = a \exp\left(-\frac{1}{N_{0i}} \int_{t-\Delta t}^{t} u_{\text{ш}i}^2 \, dt\right), \quad (18.128)$$

where a -- some constant. Therefore

$$W_i(y^{(i)} - u_{ci}) = a \exp\left\{-\frac{1}{N_{0i}} \int_{t-\Delta t}^{t} [y^{(i)}(t) - u_{ci}(t, x)]^2 dt\right\}, \quad (18.129)$$

and expression (18.127) takes the form

$$P_x(\vec{y}) = a \exp\left\{-\int_{t-\Delta t}^{t} \sum_{i=1}^{m} \frac{1}{N_{0i}} [y^{(i)}(t) - u_{ci}(t, x)]^2 dt\right\}. \quad (18.130)$$

Here, from formula (18.126) we have /355

$$\int_{t-\Delta t}^{t} Q(\vec{y}, x, t) \, dt = -\ln a + \int_{t-\Delta t}^{t} \sum_{i=1}^{m} \frac{1}{N_{0i}} [y^{(i)}(t) - u_{ci}(t, x)]^2 dt. \quad (18.131)$$

Since only the second derivative of function $Q(\vec{y}, x, t)$ with respect to x will be included in expression (18.125), then it is possible to disregard the term ln a not depending on x in (18.131) and to assume

$$Q(\vec{y}, x, t) = \sum_{i=1}^{m} \frac{1}{N_{0i}} [y^{(i)}(t) - u_{ci}(t, x)]^2. \quad (18.132)$$

Following substitution of this expression into formula (18.125) and appropriate transforms, we obtain (considering relationship (18.59) and the smallness of magnitude $\varepsilon = x - \gamma$):

$$
g_{m\text{э}} = \frac{1}{\sum_{i=1}^{m} \frac{1}{N_{0i}} B_i}, \quad \text{where} \quad B_i = \overline{\left[\frac{\partial u_{ci}(t,\ x)}{\partial x}\right]^2}. \tag{18.133}
$$

When m = 1, formula (18.133), as could be expected, coincides with corresponding formula (18.31) for a single-channel system.

Thus, all conclusions drawn for a single-channel index remain valid (within the framework of the same assumptions) for a multichannel index as well, the only difference being that potential spectral density $g_{m\text{э}}$ is determined from formula (18.133), rather than from formula (18.31).

The magnitude of the root-mean-square measurement error is computed with the aid of the equivalent circuit depicted in Figure 18.3c or from formula (18.13).

As noted in § 18.2, in a number of cases, (in automatic angular coordinate target tracking systems, for example), the oscillation at measurement system input has the form

$$
y(t) = u_c(t,\ \varepsilon) + u_m(t), \quad \text{where} \quad \varepsilon(t) = x(t) - \gamma(t) \tag{18.134}
$$

-- error signal equalling true parameter value deviation from the measured value.

Here, all formulas obtained for potential single-channel measurement accuracy remain valid if x(t) is replaced by ϵ (t) when $g_{m\text{э}}$ is computed. It is possible analogously to demonstrate that, in multichannel indices, if oscillations /356
at measuring channel input, as opposed to expression (18.123), have the form

$$\left.\begin{aligned} y^{(1)}(t) &= u_{c1}(t, \varepsilon) + u_{m1}(t), \\ &\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots \\ y^{(m)}(t) &= u_{cm}(t, \varepsilon) + u_{mm}(t), \end{aligned}\right\} \tag{18.135}$$

then the above results remain valid if ϵ replaces x in expression (18.133), i. e., if you assume

$$g_{m0} = \frac{1}{\sum_{i=1}^{m} \frac{1}{N_{0i}} B_i}, \tag{18.136}$$

where

$$B_i = \overline{\left[\frac{\partial u_{ci}(t, \varepsilon)}{\partial \varepsilon}\right]^2}. \tag{18.137}$$

In the particular case when the spectral densities of the noise at all measurement channels are identical, i. e.,

$$N_{01} = N_{02} = \ldots = N_{0m} = N_0,$$

formula (18.136) takes the form

$$g_{m0} = \frac{N_0}{\sum_{i=1}^{m} B_i}. \tag{18.138}$$

In an even more particular case when

$$B_1 = B_2 = \ldots = B_m = B,$$

the result is

$$\left.\begin{aligned} g_{m0} &= \frac{N_0}{mB}, \\ \text{where} \quad B &= \overline{\left[\frac{\partial u_c(t, \varepsilon)}{\partial \varepsilon}\right]^2}. \end{aligned}\right\} \tag{18.139}$$

As our illustration, we will use the resultant general relationships to compute potential multichannel direction finder accuracy.

For a two-channel direction finder with two antennas forming an amplitude equisignal zone, voltages of the signals at the input of measuring channels equal, respectively

$$\left.\begin{aligned} y^{(1)}(t) &= u_{c1}(t, \varepsilon) + u_{m1}(t) \\ y^{(2)}(t) &= u_{c2}(t, \varepsilon) + u_{m2}(t), \end{aligned}\right\} \tag{18.140}$$

where /357

$$\left.\begin{aligned} u_{c1}(t, \varepsilon) &= [1 + \mu\varepsilon(t)]\, b(t), \\ u_{c2}(t, \varepsilon) &= [1 - \mu\varepsilon(t)]\, b(t), \end{aligned}\right\} \tag{18.141}$$

$\varepsilon(t)$ -- angular deviation of the target the bearing is taken on from the axis of the equisignal zone; μ -- coefficient depending on the shape of the radiation pattern of the antennas and on the magnitude of the angular deviation of the main pattern maxima from the axis of the equisignal zone; b(t) -- signal carrier oscillation.

Assuming the noise intensity of both channels is identical, in accordance with formulas (18.137) and (18.138) we obtain

$$g_{m\varepsilon} = \frac{N_0}{2\mu^2 P_{cp}}, \tag{18.142}$$

where P_{cp} -- specific mean power of the signal b(t) carrier oscillation.

Given an m-channel direction finder with a phased array antenna system, it will comprise m equidistant antenna elements $H_1, H_2, \ldots, H_{m/2}, H_{-1}, H_{-2}, \ldots, H_{-m/2}$ (Figure 18.13), whose voltages are supplied correspondingly to m individual amplifier stages and then to the processing system, which extracts error signal $\varepsilon(t)$ from the combination of received oscillations, i. e., the angular deviation of the target the bearing is taken on U from axis z of the system equisignal zone. Here, the combination of oscillations at measurement system input may be written in the form

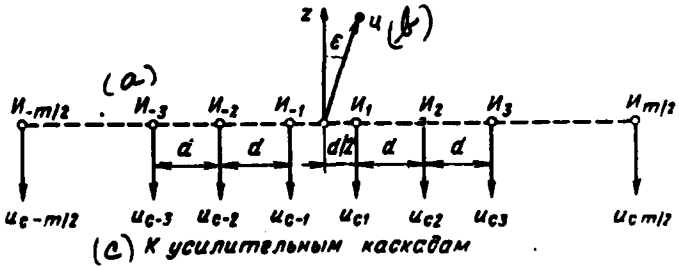


Figure 18.13. (a) -- I [Antenna elements]; (b) -- Ts [Target]; (c) -- To amplifier stages.

$$\left.\begin{array}{l} y^{(1)}(t)=u_{c\,1}(t,\varepsilon)+u_{ш\,1}(t), \\ \dots\dots\dots\dots\dots\dots \\ y^{(m/2)}(t)=u_{c\,m/2}(t,\varepsilon)+u_{ш\,m/2}(t), \\ y^{(-1)}(t)=u_{c\,-1}(t,\varepsilon)+u_{ш\,-1}(t), \\ \dots\dots\dots\dots\dots\dots \\ y^{(-m/2)}(t)=u_{c\,-m/2}(t,\varepsilon)+u_{ш\,-m/2}(t), \end{array}\right\} \tag{18.143}$$

where $u_{ш\pm k}$ -- noise at $\pm$ k-th channel input.

In future, we will assume for simplicity that all antenna elements and individual amplifier stages are identical. Then, spectral densities of the noise at the input of each channel are identical and equal N_0, while signal voltages, given slight angular deviations ϵ, have the form

$$\left.\begin{array}{l} u_{c\pm 1}(t,\varepsilon)=U(t)\cos[\omega t+\varphi(t)\pm h\varepsilon(t)], \\ \dots\dots\dots\dots\dots\dots \\ u_{c\pm k}(t,\varepsilon)=U(t)\cos[\omega t+\varphi(t)\pm(2k-1)h\varepsilon(t)], \\ \dots\dots\dots\dots\dots\dots \\ u_{c\pm m/2}(t,\varepsilon)=U(t)\cos[\omega t+\varphi(\,)\pm(m-1)h\varepsilon(t)]. \end{array}\right\} \tag{18.144}$$

where $h=\frac{\pi d}{\lambda};\ \lambda=\frac{2\pi c}{\omega}$. /358

Expressions (18.144) should be substituted into formulas (18.137) and (18.138) to determine potential measurement accuracy. Here, the result is

$$g_{ш\,\varepsilon}=\frac{N_0}{2\sum\limits_{k=1}^{m/2}B_k}, \tag{18.145}$$

where

$$B_k = \overline{\left[\frac{\partial u_{c\pm k}(t, \varepsilon)}{\partial \varepsilon}\right]^2} = (2k-1)^2 \frac{h^2}{2} \overline{U^2(t)}.$$

Considering that

$$P_{cp} = \frac{\overline{U^2(t)}}{2} \tag{18.146}$$

is specific mean signal power at the output of each emitter, from (18.145) we will obtain

$$g_{m\,*} = \frac{N_0}{2h^2 P_{cp} \sum_{k=1}^{m/2} (2k-1)^2}.$$

But

$$\sum_{k=1}^{m/2} (2k-1)^2 = \frac{m^3-m}{6};$$

therefore, finally we obtain

$$g_{m\,*} = \frac{3N_0}{h^2 P_{cp} (m^3-m)}. \tag{18.147}$$

18.6 Concluding Comments

Five basic assumptions forming the foundation of the theory of optimum nonlinear filtration were pointed out in the beginning of this chapter (§ 18.1). We will examine how much and in what sense these assumptions may reduce this theory's field of use.

The first assumption is that the law of message x(t) distribution is normal.

If this law differs significantly from the norm, then this theory in the general case will be imprecise; in particular, nonlinear smoothing networks (instead of linear filter Φ_1) may be required. However, as pointed out above, if, in accordance with problem conditions, smoothing networks fall into the class of linear networks, then all basic results of the theory are valid for an abnormal

law of message distribution (if the remaining assumptions, less the first one, remain valid).

In addition, as will be demonstrated in the next chapter, the higher the message fidelity requirements, the less message characteristics impact upon message fidelity. Therefore, one may consider that the first assumption does not /359 overly restrict this theory's field of use.

The <u>second assumption</u> is that message x(t) changes slowly in comparison to additive noise $u_m(t)$ and compared to parasitic signal parameters changing over time, i. e., compared to signal fluctuations.

The first part of this assumption [x(t) slowness compared to $u_m(t)$] is satisfied in a majority of practical cases. The slowness of x(t) compared with signal fluctuations (if they exist) occurs in far from all practical cases.

For example, in radar problems when target motion parameters are messages, the correlation time of these messages usually is measured in unities of seconds or more, while the correlation time of signal fluctuations are tenths of seconds or less. Therefore, the second assumption usually is satisfied to a lesser degree than is the first assumption. However, in a case of reception of voice or, moreover, television messages, the rate of message change is not less, but rather is significantly greater, than the signal fluctuation rate (fading). Therefore, this theory may be considered valid for such messages only in the assumption that signal fluctuations are completely absent. Such an assumption contradicts reality, but makes it possible to obtain the potential (<u>theoretically</u>-permissible threshold) message fidelity value (since the fact that a real signal has fluctuations may only lower this fidelity). Consequently, for "fast" messages, this theory makes it impossible to find optimum receiver structure and applies only to potential message fidelity estimation.

The <u>third assumption</u> (concerning high message fidelity) in a majority of practical cases may be considered satisfied, at least in the first approximation.

The <u>fourth assumption</u> is that modulation of signal $u_c(t; x; \alpha_1, ..., \alpha_m)$ by message x(t) is direct.

This assumption is invalid for several important cases, frequency modulation in particular. However, as shown in § 18.4, it is possible to generalize the theory for cases of indirect modulation as well when determining potential message fidelity.

The fifth assumption (concerning the stationary nature of signal fluctuations) in a majority of practical problems may be considered valid, at least in the first approximation. In addition, the Gauss approximation method remains valid also when this assumption is removed, although computations and results here are complicated considerably (see [127], for example).

It follows from what has been said that, although the assumptions made considerably reduce the generality of the theory examined in this chapter, its domain remains very broad, especially when determining potential message fidelity.

CHAPTER NINETEEN /360

IMPACT OF OPTIMIZATIONS AND A PRIORI MESSAGE DISTRIBUTIONS ON OPTIMUM RECEIVER STRUCTURE AND PROPERTIES

19.1 General Comments

In accordance with the theory presented above, finding the structure of the optimum receiver and determining its noise immunity require precise knowledge of the a priori signal and noise distribution* and selection of the corresponding optimization, i. e., the loss function $l(x, \gamma)$ type. However, significant difficulties arise here since a priori distributions usually are not precisely known and unambiguous selection of the specific loss function type is not always possible.

Initially, we will examine the a priori signal and noise distribution problem.

The a priori distribution of the basic noise type--internal receiver noise, is precisely known. The a priori distribution of most other noise of natural origin (atmospherics, background reflections, and so on) also usually is known or may be approximated from preliminary perennial research. On the other hand,

*A priori data may be required also for selection of anticipated signal strength and of other characteristic problem parameters.

the a priori distribution of organized noise turns out to be unknown completely or to a signifcant degree.

A priori signal distribution is determined from usable message distribution P(x) and parasitic parameter distribution $P(\alpha_1, \ldots, \alpha_m)$. (Usually, one may assume that message x and the parasitic parameters mutually are statistically independent).

In most cases, distribution $P(\alpha_1, \ldots, \alpha_m)$ is know, at least approximately. For example, it is known that, when receiving a signal reflected from an aircraft, the initial phase has a uniform distribution, while it is possible in the first approximation to assume that the amplitude is distributed in accordance with Rayleigh's law. Here, by virtue of the accumulation of experimental data, the law of amplitude distribution will become known with ever-greater accuracy.

A priori message P(x) distribution in several technical cases is known or may be determined with sufficient accuracy from experimental material, which already is available or which may be accumulated in the near future. This occurs, for instance, in civilian radio communications and during radio and TV broadcasts. However, in many cases, the a priori message distribution is unknown and, in principle, may not be found due to the absence of preliminary similar situations, know-how from which could be put to use. This happens, for example, in /361 military radar and during mensuration involving the study of new physical phenomena.

Thus, in a majority of cases, a priori signal and noise distributions are not precisely known, with this applying in particular to a priori message distributions. However, it turns out that, in many important cases, optimum receiver structure and properties essentially will not depend on a priori distributions of messages and certain parasitic signal parameters. Thus, for instance, it was pointed out above many times that, given noise in the form of additive normal white noise and a high signal-to-noise ratio, the noise immunity of the optimum receiver, given a signal with an equiprobable random initial phase, turns out to be identical to that for a precisely-known signal. This signifies that, under the conditions stipulated, the law of signal initial phase φ distribution does not impact on optimum receiver properties.

It will be demonstrated below that, under these conditions (given noise in

the form of additive normal white noise and high signal-to-noise ratio), the type of a priori message P(x) distribution also essentially does not impact upon optimum receiver structure and properties, at least in a case of individual analog message value reception.

We now will examine the problem of selecting a specific type of loss function $l(x, \gamma)$. It follows from material in the preceding chapter that, when receiving discrete messages, loss function selection boils down to selection of the weights of various error types (false alarm and signal miss weights, for example). In several cases, selection of these weights does not present difficulties. Thus, for instance, during transmission of an alphabet with equiprobable letters lacking a statistical link among them, it usually is possible to assume that the weight of all errors is identical, i. e., to consider that it is identically-dangerous to receive any other letter of the alphabet instead of a given letter. However, in many cases, the physical crux of the problem does not make it possible to select error weight so that there is confidence that this weight selection is the best or the only possible one.

In a case of simple reproduction of analog messages, as already noted above, a quadratic loss function of the following type usually is selected as the loss function

$$l(x, \gamma) = (\gamma - x)^2.$$

However, in many cases, it is far from evident that such a loss function is the best or the only correct one. For instance, the advantages or disadvantages of this loss function compared to a function of the type $l(x, \gamma) = |\gamma - x|$ and so on may be unclear.

Thus, there is considerable ambiguity in selection of the type of a priori distributions and loss functions in a majority of cases during the design of an optimum receiver and evaluation of its quality. It also is evident that selection of a particular loss function type in turn is based on certain a priori data. Thus, for example, if it is known a priori that an alphabet is the message, /362 then this provides the basis to select identical weights for various errors. If it is known that music or speech is the message, a quadratic loss function

is advisable in both cases. However, a priori data never may be comprehensive and a considerable degree of ambiguity (randomness) remains in selection of the loss function type, even in the best cases. Therefore, it is very important to clarify the degree to which this ambiguity may impact upon optimum receiver structure and properties. Study of the problem is one of the basic tasks of the contemporary theory of optimum reception methods.

It is shown in succeeding sections that, given noise in the form of additive normal white noise (and not only that type) and high requirements for message fidelity (reliability), the structure of the optimum receiver and its properties essentially will not depend on the type of a priori message distribution and on the optimization selected for a broad class of these distributions and criteria. Therefore, in the quest for the optimum receiver and study of its properties, it is permissible in many cases to begin with the simplest (uniform) law of message distribution and from one of the simplest optimizations.

19.2 Certain Properties of Distribution $P_y(x)$ During Reception of Individual Analog Message Values

The optimum receiver must reproduce message x in the best-possible manner based on analysis of signal-plus-noise realization y(t). For a given y(t), all usable information on message x will be contained in distribution $P_y(x)$; therefore, optimum receiver structure and properties are determined primarily by the distribution $P_y(x)$ for any optimization.

Thus, for example, in the case of the maximum inverse probability criterion [condition (17.35)], the task of the optimum receiver is simply to find the distribution $P_y(x)$ maximum (with respect to x) and, using the generalized minimum average risk criterion [condition (17.32)], receiver structure and properties also are determined primarily by distribution $P_y(x)$ type.

Therefore, we must first explain basic distribution $P_y(x)$ properties in order to analyze the influence of a priori distribution P(x) and loss function $l(x, \gamma)$ type.

We will examine these properties relative to a case of simple reproduction of individual analog message values, when distribution $P_y(x)$ is a unidimensional

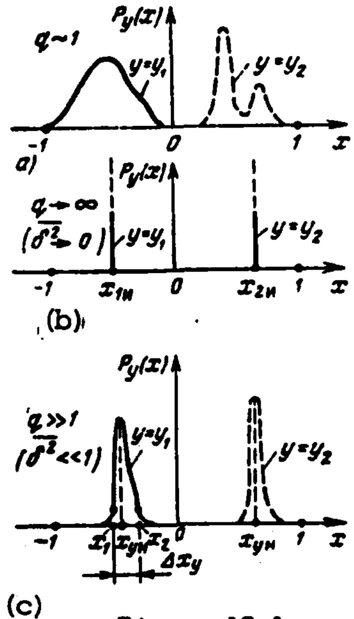


Figure 19.1

probability density and therefore easily is depicted graphically (Figure 19.1). Here, we will assume, as was the case in Parts I and II of the book, that the message is normalized so that

$$-1 \leqslant x \leqslant 1. \tag{19.1}$$

This normalization does not disrupt the generality of the analysis, /363
while, for any actual message τ change (finite) bounds $\tau_{мин} - \tau_{макс}$, it is possible, having assumed

$$\left.\begin{aligned} &\tau = \tau_0 + \tau_1 x, \\ \text{where} \quad & \\ &\tau_0 = \frac{\tau_{макс} + \tau_{мин}}{2}; \quad \tau_1 = \frac{\tau_{макс} - \tau_{мин}}{2}, \end{aligned}\right\} \tag{19.2}$$

to obtain normalized message x change bounds equalling $\pm$ 1. Here, τ_0 and τ_1 are known magnitudes and transform (19.2) denotes only coordinate origin displacement and introduction of a specfic scale with respect to the X-axis.

The type of curve $P_y(x)$ for different realizations y(t) ($y = y_1$ and $y = y_2$

in Figure 19.1, for example), in the general case may vary, while the area is determined by normality condition

$$\int_{-1}^{1} P_y(x)\,dx = 1, \qquad (19.3)$$

i. e., always equals unity.

We will assume that, when there is no noise, message x will be reproduced precisely, i. e., message reproduction errors are linked exclusively with noise activity. This signifies, in particular, that, if the signal has parasitic random parameters $(\alpha_1, \ldots, \alpha_m)$, then the nature of these parameters is such that it does not hinder, in principle, accurate message mensuration when noise is absent. Such parasitic parameters in future for brevity will be called safe parasitic parameters.

For instance, let

$$u_0(t) = a_0(1+x)\cos(\omega t + \varphi), \qquad (19.4)$$

where φ -- equally-probable initial phase; while a_0 and ω -- precisely-known parameters.

Evidently, in this case, parasitic parameter φ is safe since, when there is no noise, it is possible through amplitude detection to extract envelope $a_0(1 + x)$ and, consequently, in principle to measure message x accurately.

Since the message will be reproduced accurately when noise is absent, distribution $P_y(x)$ here has the form of delta-function $\delta(x - x_H)$, where x_H — true message value (Figure 19.1b).

Since reproduction error in this example equals zero, we will call this /364 a case of zero error and designate*

$$\overline{\delta^2} = 0. \qquad (19.5)$$

*In expression (19.5), δ denotes error, rather than a delta-function.

Consequently, distribution $P_y(x)$ has the form of a delta-function in the case of zero error.

In a real case, i. e., given fluctuating noise, the error does not equal zero. However, given known conditions, it is sufficiently slight, i. e.,

$$\overline{\delta^2} \ll 1. \tag{19.6}$$

It follows from what has been stated above that, when $\overline{\delta^2} \to 0$, distribution $P_y(x)$ will strive towards a delta-function. Therefore, given a finite but slight-enough error, this distribution must have a form close to a delta-function (Figure 19.1c), i. e., curve $P_y(x)$ in this case has a significant value only within very-narrow interval Δx_y around the most-probable value $x_{y\text{и}}$. Here, index y denotes that magnitudes Δx_y and $x_{y\text{и}}$ in the general case will depend on realization y(t) type. However, due to the error decrease, magnitude Δx_y asymptotically will strive towards zero essentially for all possible realizations y(t) (since, otherwise, error $\overline{\delta^2}$ would not strive towards zero).

Since error $\overline{\delta^2}$ asymptotically will strive towards zero given an unlimited rise in signal-to-noise ratio q, instead of (19.5) and (19.6), it is possible to assume, respectively

$$q \to \infty \tag{19.7}$$

and

$$q \gg 1. \tag{19.8}$$

In the first case (where $q \to \infty$), distribution $P_y(x)$ is converted into a delta-function, while, in the second ($q \gg 1$), it has significant values only within very slight interval Δx_y.

It is possible to use the results presented in Parts II and III for noise in the form of additive normal white noise, valid for a high signal-to-noise ratio ($q \gg 1$), to illustrate these general postulations about the function $P_y(x)$ type.

For a precisely-known signal, distribution $P_y(x)$ is described by formula (6.12):

$$P_y(x)=\sqrt{\frac{b}{\pi}}\,e^{-b\,(x-x_{y\,н})^2},\tag{6.12}$$

while the mean-square error equals

$$\overline{\delta^2}=\frac{1}{2b},\tag{6.17}$$

where /365

$$b=\frac{T}{N_0}\left(\left[\frac{\partial u_x(t)}{\partial x}\right]^2_{x_{y\,н}}\right)_T.$$

In a case of amplitude modulation

$$b=\frac{m^2Q_0}{N_0}=q_0m^2.\tag{6.22}$$

For frequency or pulse-position modulation, as follows from (6.17), (6.26b), and (6.40), the result is

$$b=\frac{(\Omega T)^2}{12}\cdot\frac{Q}{N_0}=\frac{(\Omega T)^2}{12}q.\tag{19.9}$$

It follows from these relationships that, where $\overline{\delta^2}\to 0$ or $q\longrightarrow\infty$, distribution $P_y(x)$ actually will strive towards a delta-function. In addition, in the cases

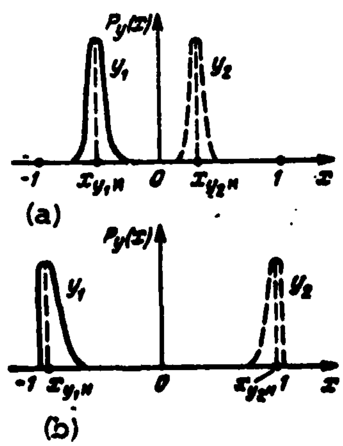


Figure 19.2

examined, distribution $P_y(x)$ has several additional special features, which are important for further analysis:

a) distribution $P_y(x)$ is a single-humped curve symmetrical relative to most-probable value x_{yn} (Figure 19.2a);

b) the type of curve $P_y(x)$ for various realizations y(t) remains unchanged. When y changes, only a displacement along the X-axis of the position of the curve's maximum caused by the dependence of y on most-probable value x_{yn} will occur [in the cases examined, evidently $x_{yn} \approx x_n$, while true value x_n may be anything ranging from —1 to 1].

Postulations a) and b) are valid for all realizations y, with the exception of those which correspond to values x_{yn}, very close to $\pm$ 1.

If x_{yn} is close to $\pm$ (Figure 19.2b), distribution $P_y(x)$ will become asymmetrical and will depend on y. Such a distribution $P_y(x)$ distortion occurs because the probability of message x values falling into the $\pm$ 1 range must equal zero. However, if $q \gg 1$ (or error $\overline{\delta^2}$ is very slight), then curve $P_y(x)$ has a very narrow peak and its significant distortion may occur only for x_{yn} values very close to the boundaries. The probability of such values is very slight. Actually, in the case being examined of very-slight errors for the overwhelming majority of realizations y(t), most-probable message value x_{yn} must be very close to its true value x_n. Therefore, if a priori distribution P(x) does not have sharp peaks around the boundary values ($x \approx \pm 1$), then the probability of the appearance of message value x_n very close to one of the boundary values is very slight. Consequently, the probability that most-probable message value x_{yn} will turn out to be close to one of the boundaries also is very slight.

Thus, with the exception of special cases [in which a priori distribution P(x) has sharp maxima around boundary values ($x = \pm 1$)], it is possible to assume that properties a) and b) occur, i. e., distribution $P_y(x)$ is a single-humped /366 curve symmetrical relative to most-probable value x_{yn} and, this curve, (without distortions) will displace along the X-axis only when y changes (Figure 19.2a).

Kotel'nikov obtained formulas (6.12) and (6.17) and drew conclusions from them for precisely-known signals. However, it was demonstrated in Part III that,

given a high signal-to-noise ratio, they remain valid also for random initial phase signals.

So, distribution $P_y(x)$ is a symmetrical single-humped curve essentially for all y, given additive noise in the form of normal white noise and a high signal-to-noise ratio ($q \gg 1$).

At first glance, it may be demonstrated that, since, where $q \longrightarrow \infty$, distribution $P_y(x)$ will strive towards a delta-function, when q increases, it must strive towards a single-humped symmetrical curve for any fluctuating noise, but not only in the case of normal white noise. However, in actuality, given some forms of fluctuating noise, distribution $P_y(x)$ may remain asymmetrical or not single-humped for a signal-to-noise ratio as high as desired (but finite).

The proof of this postulation was provided in [125]. It was demonstrated here that, given a high signal-to-noise ratio, distribution $P_y(x)$ may be (essentially for all y) a single-humped curve symmeterical relative to the most-probable value, not only for noise in the form of normal white noise, but for some other types of noise as well.

These conclusions are of great significance when analyzing the influence of optimizations, as will be shown in the next section.

19.3 Impact of A Priori Message Distributions and Optimizations During Reception of Individual Analog Message Values

1. Impact of A Priori Message $P_y(x)$ Distribution

In the general case, distribution $P_y(x)$ will depend on a priori message P(x) distribution since

$$P_y(x) = k_y P(x) P_x(y), \tag{19.10}$$

where k_{y1} -- coefficient determined from normality condition /367

$$\int_{-1}^{1} P_y(x)\, dx = 1. \tag{19.11}$$

Let the signal-to-noise ratio be so high that distribution $P_y(x)$ has a significant value only within the range of a very-narrow region with width Δx_y (Figure 19.1c). We will clarify what conditions distribution $P(x)$ must meet so its form essentially will not impact upon the shape of curve $P_y(x)$.

In this case, likelihood function $P_x(y)$ in accordance with (19.10) must have, accurate to some constant factor, a shape exactly like that of $P_y(x)$. Therefore,

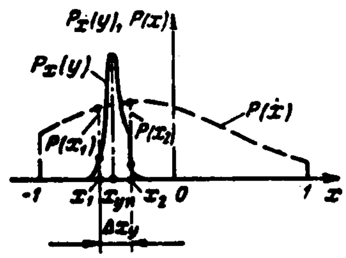


Figure 19.3

function $P_y(x)$ depicted in Figure 19.1c in this case thus will be similar in shape to likelihood function $P_x(y)$ depicted in Figure 19.3. A priori distribution $P(x)$ is shown by the dotted line in this figure.

It is easy to become convinced from comparison of Figure 19.1c and 19.3 and considering relationship (19.10) that curve $P_y(x)$ shape will not depend on the distribution $P(x)$ type if function $P(x)$ is sufficiently-smooth in the sense that, for any functions $P(x)$, condition $\frac{|P(x+\Delta x_y) - P(x)|}{P(x)} \leqslant \varepsilon$ is satisfied for all x ranging from —1 to 1, (19.12), where ϵ — positive magnitude significantly less than unity, such as $\epsilon = 0.1$.

In other words, any possible a priori distribution must be so smooth that, in very-slight interval Δx_y, function $P(x)$ changes slightly, not more than $\pm$ 10%, for example.

When the signal-to-noise ratio rises, magnitude Δx_y included in condition (19.12) asymptotically will strive towards zero. Therefore, the greater the signal-to-noise ratio, the broader the class of a priori distributions $P(x)$ for

which this condition is met. Consequently, given a sufficiently-high signal-to-noise ratio, function $P_y(x)$, and, hence, optimum receiver structure and properties, will not depend on the a priori distribution $P(x)$ type for a very-broad class of these distributions.

But, the greater the minimum signal-to-noise ratio value, the higher the message fidelity requirements. Therefore, it is possible to formulate the aforementioned conclusion also in the following manner:

Given sufficiently-high message fidelity requirements (slight $\overline{\delta^2}$), optimum receiver structure and its properties will not depend on the type of a priori /368 message distribution for a very-broad class of these distributions. Hence, it follows that, given the aforementioned conditions, during optimum receiver design and analysis of its properties, it is possible to start from the simplest a priori message distribution type, i. e., to assume

$$P(x) = \text{const.}$$

2. Optimization Impact

We will demonstrate that, given high message fidelity requirements (for a high signal-to-noise ratio), a broad class of optimizations exists, which leads to identical optimum receiver structure and properties.

For proof, we will begin from the generalized statistical criterion presented in the preceding chapter--the minimum average risk R criterion:

Average risk R is determined from formulas (17.30) and (17.31), namely

$$R=\int_{A_y} P(y)\, R_y\, dy, \tag{17.30}$$

where

$$R_y=\int_{A_x} P_y(x)\, I\,[x,\ \Gamma(y)]\, dx. \tag{17.31}$$

As indicated above, finding the optimum receiver structure will boil down

from the mathematical point of view to finding that type of operator $\Gamma(y)$ in which magnitude R is minimum. Various optimizations correspond here to different loss function $l(x, \gamma)$ types.

We will begin with analysis of expression (17.31). We will assume that, for any y, function $P_y(x)$ is a single-humped curve symmetrical relative to the most-probable value.

As was demonstrated above, such an assumption is valid for a high signal-to-noise ratio for normal white noise and some other noise types. On the other hand, it is possible to show [125] that, for some noise types, curve $P_y(x)$ remains signficantly asymmetrical or not single-humped for a signal-to-noise ratio as high as desired (but finite). Thus, our assumption that distribution $P_y(x)$ is, for all y, a single-humped curve symmetrical relative to x_{yn} is valid (given a high signal-to-noise ratio) not always, but, in any case, valid for normal white noise and not only for that.

In future, we will consider this assumption valid.

Let the receiver be built in accordance with the maximum inverse probability density criterion, i. e., it supplies always that value x which is the most /369 probable for a given y. This signifies that, in such a receiver, this always is the case

$$\Gamma(y) = x_{y\,n}.$$

Conditional risk R_y in such a receiver equals

$$R_y = \int_{A_x} P_y(x)\, l(x,\ x_{y\,n})\, dx.$$

We will explain which loss function $l(x, \gamma)$ types apply for this receiver operating principle to be optimum, i. e., provides the minimum R_y value for any y.

Function $l(x, \gamma)$ type will depend on the optimization selected. However, for rational criteria, this function usually has the following general properties:

1. $$l(x, \gamma) \geqslant 0. \qquad (19.13)$$

This denotes that any receiver error is considered a loss (deficit).

2. $$I(x, x) = 0 \qquad (19.14)$$

i. e., the loss corresponding to a precise decision $(\gamma = x)$ equals zero.

3. Function $l(x, \gamma)$ rises monotonically with respect to both sides from point $x = \gamma$ (i. e., from point I = 0).

4. In the highly-accurate systems we are examining, when possible errors $(\gamma - x)$ must be very slight, it is most natural to consider the weights of positive and negative errors to be identical. Therefore, in such systems, the loss function selected usually is symmetrical and nondecreasing. In addition, the assumption most of the time is that the loss is a function only of the error modulus $(\gamma - x)$, i. e.,

$$l(x, \gamma) = l(x - \gamma) = l(\gamma - x), \qquad (19.15)$$

where $l(\gamma - x)$ — nondecreasing function of error $(\gamma - x)$.

Functions of the type $(\gamma - x)^2$ and $|\gamma - x|$, corresponding to well-known criteria of minimum root-mean-square error and minimum error modulus, are examples of loss functions satisfying conditions (19.13)--(19.15).

We now will demonstrate that, given a single-humped distribution $P_y(x)$, the maximum inverse probability density criterion is optimum for all loss function $l(x, \gamma)$ types satisfying conditions (19.13)--(19.15).

A symmetrical single-humped distribution $P_y(x)$ and two postulations of (19.15)-type loss function $l(x, \gamma)$ are depicted in Figure 19.4a. The function $l(x, \gamma)$ postulation in a receiver operating on the maximum inverse probability principle is depicted by a solid line. Here, $\gamma = x_{ym}$. The function $l(x, \gamma)$ postulation in a receiver operating on some other principle is depicted by the dotted line.

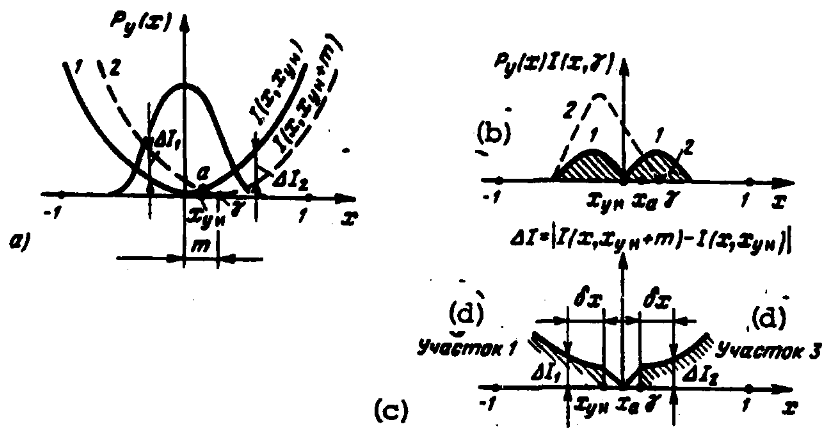


Figure 19.4. (d) -- Section.

In this case, $\gamma = x_{y\text{н}} + m(y)$, where magnitude m(y) characterizes the deviation from principle $\gamma = x_{y\text{н}}$.

Values of a derivative of functions $P_y(x)$ and $I(x, \gamma)$ are presented /370 in Figure 19.4b. The areas formed by curves 1 and 2 equal the values of conditional risk R_y for functions $I(x, x_{y\text{н}})$ and $I(x, x_{y\text{н}} + m)$, respectively [see formula (17.31)].

We will show that a larger area corresponds to curve 2 than to curve 1. For proof, we will compare the areas individually for the following three sections:

1. Areas located where $x \leqslant x_{y\text{н}}$.

2. Areas located where $x = x_{y\text{н}} \div \gamma$.

3. Areas located where $x > \gamma$.

In the first section, the area corresponding to curve 2 always is greater by some positive magnitude ΔQ_1 since, in this section, curve 2 (Figure 19.4a) is located above curve 1.

In the third section, the area corresponding to curve 2 always is less by

some positive value ΔQ_3 since, here, curve 2 (Figure 19.4a) is below curve 1 everywhere. But, $\Delta Q_1 > \Delta Q_3$ always is the case. Actually, it is evident from Figure 19.4a and 19.4c that, in the first and third sections, the dependence of the difference of ordinates ΔI of curves 1 and 2 on x is symmetrical relative to value $x = x_a$. Therefore, two differences ΔI_1 and ΔI_2, taken equal distances δx from the boundaries of the first and third sections (Figure 19.4a and 19.4c), are equal among themselves, but function $P_y(x)$ at the point corresponding to ΔI_1 is greater than at the point where the ordinate difference equals ΔI_2 (Figure 19.4c). This occurs for any value δx within the first and third sections. Therefore, it is evident that a change in the area formed by curve $P_y(x)\, I(x, \gamma)$ causing a shift from function $I(x, x_{ym})$ to function $I(x, x_{ym}+m)$ will, in the first section, always be greater (with respect to absolute magnitude) than in the third section, i. e., actually $\Delta Q_1 > \Delta Q_3$.

In the second section $(x = x_{ym} \div \gamma)$, the area formed by curve 2 is greater by positive magnitude ΔQ_2 than the area formed by curve 1. Actually, curves 1 and 2 (Figure 19.4a) in this section are located symmetrically relative to the mean abcissa x_a, but the curve 2 greater ordinate values correspond to greater function $P_y(x)$ ordinate values. Therefore, following multiplication of the corresponding ordinates, the area formed by curve $P_y(x)\, I(x, \gamma)$, in the case /371 of curve 2, ends up greater. Thus, the overall area increment equals

$$\Delta Q_1 + \Delta Q_2 - \Delta Q_3 > 0,$$

and magnitude R_y rises during the shift from $m = 0$ to $m \neq 0$.

Consequently, conditional risk R_y is minimal where $m = 0$, i. e., when decision γ selected each time equals most-probable value x_{ym}.

Thus, it is demonstrated that, given the aforementioned assumptions about function $P_y(x)$ and $I(x, \gamma)$ type, a receiver operating in accordance with the maximum inverse probability density principle $(\gamma = x_{ym})$ provides minimum conditional risk R_y for any y and, consequently, minimum average risk as well [see formula (17.30)].

Hence, it follows that, in those cases when distribution $P_y(x)$ may be considered a single-humped curve symmetrical relative to the most-probable value for all

y, a receiver operating in accordance with the maximum inverse probability density principle is optimum (i. e., which insures minimum average risk) for all loss functions satisfying conditions (19.13)--19.15). In particular, this occurs given a high signal-to-noise ratio for normal white noise and not only for such noise.

It is easy to demonstrate (see [125], for example) that, in those cases when distribution $P_y(x)$ or loss function $l(x, \gamma)$ have no symmetry, a receiver operating on the maximum inverse probability density principle is not optimum (i. e., does not provide minimum average risk), while optimum receiver structure will depend on specific loss function type.

Based on the aforementioned analysis, it is possible to draw the following conclusions valid for precisely-known signals or those with safe random parameters:

1. Given high message fidelity, optimum receiver structure and properties essentially will not depend on the type of a priori distribution $P(x)$. Therefore, when seeking the optimum receiver and evaluating its properties, given the aforementioned conditions, it is possible to assume that distribution $P(x)$ is the simplest, i. e., uniform.

2. If distribution $P_y(x)$ may be considered, for all y, a single-humped curve symmetrical relative to most-probable value $x_{вм}$ (which occurs, in particular, for normal white noise and some other noise, given a high signal-to-noise ratio), then optimum receiver structure will not depend on the accepted optimization during the search for a broad class of these criteria. All criteria based on average risk minimization, for which the loss functions satisfy conditions (19.13)--(19.15), fall into this class.

These conditions are sufficiently general and the corresponding optimization class is sufficiently broad--it includes, in particular, criteria of maximum inverse probability density, maximum likelihood, minimum mean value of any even degree of error, minimum mean value of any odd degree of the error modulus (minimum /372 of $\overline{(\gamma - x)^{2k}}$ and $\overline{|\gamma - x|^{2k-1}}$, respectively), and others.

3. It follows from point 2 that, when seeking the optimum receiver structure, it is possible under these conditions to start from any simplest criterion which

falls into the aforementioned class, the maximum inverse probability density criterion or of the maximum root-mean-square error criterion, for example.

4. It follows from the aforementioned points that all results of the analysis of reception of individual analog message values presented in Parts II and III of this book for normal white noise, stemming from uniform a priori distribution P(x) and the maximum inverse probability density criterion, remain valid for a broad class of a priori distributions P(x) and optimizations, if the following conditions are met:

a) the signal is precisely known or has safe random parameters;

b) the signal-to-noise ratio is high enough (message fidelity is high enough).

19.4 Impact of A Priori Message Distributions and Optimizations During Discrete Message Reception

Let message x have m + 1 possible values x_0, x_1, . . ., x_m, where x_0 -- zero message corresponding to no signal.

Evidently, the greater the m, the closer the discrete message is to the character of an analog message, already examined in the preceding section. Therefore, we will restrict ourselves here to examination of another extreme case, corresponding to m = 1, i. e., to binary detection.

It was shown in Chapter 14 that, given different optimizations (Neyman-Pearson, minimum composite error probability, and others), optimum receiver structure remains unchanged--only threshold bias at receiver output changes. The magnitude of this bias for some optimizations will depend also on ratio $P(x_0)/P(x_1)$ of the a priori signal absence and presence probabilities.

Consequently, for a broad class of optimizations examined, optimum receiver structure, except for output threshold, will not depend on the optimization selected and a priori message distribution. Therefore, all that remains is to explain the extent to which receiver output threshold will depend in the general case on selected optimization and on a priori message distribution.

A study of this problem conducted in [125] made the following results possible:

1. During reception of discrete messages, as was the case for analog messages, the impact of a priori message distributions and optimizations (error weights) /373 on optimum receiver structure and properties monotonically decreases with an increase in signal-to-noise ratio, striving asymptotically towards zero.

2. The point 1 postulation is valid at least for noise in the form of additive normal white noise and a precisely-known signal.

3. If the signal has parasitic random parameters, then various cases are possible, depending on the nature of these parameters.

The point 1 postulation remains completely valid for some types of parasitic random parameters. Such random parameters may be called safe in an analogy with the case of analog messages. This includes, in particular, the random initial phase of signal rf occupation.

The point 1 postulation may be satisfied, but not fully, for other types of parasitic random parameters. This includes, in particular, the amplitude of a fluctuating signal distributed in accordance with Rayleigh's law.

4. Error weights impact upon optimum receiver structure and properties just like a priori message probabilities do.

19.5 Conclusion

The analysis presented in this chapter makes it possible to draw the following general conclusions valid for reception of discrete messages and individual analog message values.*

1. Given an increase in signal-to-noise ratio (increase in message reception accuracy or reliability), the impact of the a priori distribution P(x) type on

*These conclusions may not be extended completely to reception of oscillations (filtration). This is because, in the case of filtration, distributions P(x) and $P_y(x)$ are not uniform, but are n-dimensional; here, magnitude $n=2f_m T$ in many cases in very great and even will strive towards infinity.

optimum receiver structure and properties asymptotically will strive towards zero if the signal is precisely known or has safe random parameters.

2. In particular, initial phase φ_0 of rf occupation (constant during time of observation T) falls in the category of safe random parameters.

3. It follows from point 1 that, when designing the optimum receiver and analyzing its properties, it is possible when the signal-to-noise ratio is high to start in the first approximation from uniform a priori distribution P(x) if the signal is precisely known or has safe random parameters (random initial phase, in particular).

4. Given a high signal-to-noise ratio and noise in the form of normal white noise (and not only given such noise), a receiver operating in accordance with the maximum inverse probability $P_y(x)$ principle (maximum inverse probability density in the case of analog messages) is optimum for a broad class of optimizations if the signal is precisely known or has safe random parameters. /374

In the case of analog messages, all criteria based on average risk minimization and loss function $l(x, \gamma)$, being a nondecreasing function of error modulus $|\gamma - x|$, will fall into this class. In particular, the minimum root-mean-square error criterion will fall in this class.

5. It follows from point 4 that, when designing the optimum receiver and analyzing its properties, given a high signal-to-noise ratio, it is possible to begin in the first approximation from any simplest optimization belonging to the aforementioned class if noise has the form of normal white noise and the signal is precisely known or has safe random parameters. In particular, it is possible to begin from the maximum inverse probability density $P_y(x)$ criterion or the minimum root-mean- square error criterion.

6. Postulations 1—5 are more accurate, the higher the signal-to-noise ratio (the higher the requirement for message reception fidelity or reliability).

More-detailed analysis of each specific case is required for determination of the error magnitude in the general postulations presented for a given signal-to-noise ratio (given specific message reception reliability).

CHAPTER TWENTY /374

CERTAIN LIMITATIONS INHERENT IN STATISTICAL DECISION THEORY AND WAYS TO SURMOUNT THEM

20.1 General Comments

Statistical decision theory presented has a rather general nature and will find wide use for solution of practical problems. However, as is the case with any theory, it is based on a number of assumptions and limitations, which, in some cases, may not be satisfied. Therefore, to avoid errors, it is necessary to present these assumptions and limitations clearly. The following are the most important ones:

1. The assumption is that a priori distributions of random actions (messages, parasitic signal parameters, noise) are precisely known.

2. The assumption is that loss function $l(x, \gamma)$ is known.

3. It is assumed that the characteristics of all actions will not depend on which system is designed.

4. No limitations are placed on the system designed, other than the general requirement for physical realization (which usually means that the reaction /375

at system output must not arise prior to onset of its causal input action) and, if necessary, requirements that the system action algorithm be determinate (non-randomized).

5. System quality is evaluated only by message fidelity, i. e., other important system quality indicators (cost, complexity, weight, overall dimensions, reliability, and others) are not considered.

6. One number--<u>average</u> risk, characterizes message fidelity.

Several ways to ameliorate certain limitations already have been noted, but the overall problem, in light of its importance, requires more-detailed examination. The next chapter is devoted to the limitation noted in point 1, while the remaining limitations are examined in this chapter.

20.2 Limitations on Selection of Loss Function $I(\lambda, \gamma)$ Type

As noted in the preceding chapter, in a case of high message fidelity (validity), the structure of the system designed often is not critical where loss function type is concerned within the broad class of such functions used in practice. However, there may be cases in which it is impossible a priori to be sure that the selected loss function corresponds well enough to the physical crux of the problem and that a change in its type will not lead to a significant change in the structure and properties of the system designed. What does one do in such cases?

It is impossible to provide universal and comprehensive recommendations here. Therefore, we only will note several possible variants for surmounting this difficulty.

<u>First variant</u>. Undertake system design for not one, but several, significantly-different rational loss function types and compare the results obtained. If results turn out to be relatively close (i. e., a system which provides minimum average risk R for one loss function provides value R close to minimum for other loss functions as well), then one may consider the system design completed. Otherwise,

other system design methods not requiring provision of a specific loss function type must be used. Several of these are described below.

Second variant. Use the maximum inverse probability criterion (for discrete messages) or the maximum inverse probability density criterion (for analog messages).

However, it was demonstrated in § 17.2 that the maximum inverse probability density criterion coincides with the minimum average risk criterion, given a simple loss function type ((17.36a). Therefore, the second approach examined is equivalent to selection of a (17.36a)-type loss function.

Third variant. If not only the loss function $l(x, \gamma)$ type, but a priori message distribution P(x) also, are unknown, the maximum likelihood criterion /376 may be used [or, which is equivalent, the second variant may be used, having assumed here that distribution P(x) is uniform].

Fourth variant. During binary reception (i. e., during binary detection or recognition of two non-zero signals), when the reproduced message may have only two values, x_1 and x_2, it is convenient to characterize reception quality by two quality indicators--conditional error probabilities

$$k_1 = P_{x_1}(x_2) \quad \text{and} \quad k_2 = P_{x_2}(x_1)$$

(evidently, in a case of binary detection, x_1 and x_2 should be replaced by x_0 and x_1, respectively and this should be assumed: $k_1 = P_{x_0}(x_1) = P_{\text{лт}}$, $k_2 = P_{x_1}(x_0) = P_{\text{np}}$. Here, it is possible to seek the optimum system in accordance with the criterion

$$k_2 = \min, \text{ where } k_1 = \text{const}.$$

As will be demonstrated in Chapter 22, such a result will be obtained if you use the criterion

$$k_1 = \min, \text{ where } k_2 = \text{const}.$$

Evidently, in a case of binary detection, criterion $k_2 = \min$, where $k_1 = \text{const}$ is nothing other than a Neyman-Pearson criterion.

20.3 Assumption 3 Impact

It is assumed in optimum receiver system design that the distribution of probabilities $P(x, y) = P(x)P_x(y)$ will not depend on the type of needed decision rule $\Gamma(y)$. This is equivalent to the supposition that message, parasitic signal parameter, and noise characteristics will not depend on designed receiver structure. In actuality, this supposition may turn out to be invalid, at least for the following reasons.

First, if the actual system is built in accordance with optimum decision rule $\Gamma_{\text{кр}}(y)$ found, then this system must to some degree differ from the ideal system, due to presence of internal instabilities and noise, for example. Considering internal instabilities and noise, a system realizing some non-ideal rule $\Gamma(y)$ may turn out to be better than a system realizing rule $\Gamma_{\text{кр}}(y)$, i. e., the system design result may be incorrect.

The aforementioned idealization, in particular, led to the fact that amplifiers will not be included in the structure of all optimum receivers described in preceding chapters (Figures 4.3, 5.2, 5.3, and so on). There may be no amplification of mixture y(t). This postulation would be valid if the mixture y(t) processor [i. e., device realizing optimum algorithm $\Gamma_{\text{кр}}(y)$] would not include any additional noise. However, any actual processor will contain internal noise. Therefore, in order /377 for resultant algorithm $\Gamma_{\text{кр}}(y)$ actually to be optimum, in the general case noise action of the processor realizing this algorithm must be infinitesimally slight, i. e., processor output signal level must be high compared to processor internal noise level. This signifies that an amplifier providing requisite amplification must in the general case be connected between receiver antenna and processor.

As was demonstrated in preceding chapters, a linear filter usually is the optimum receiver input section. Here, preliminary amplification and linear filtration may be combined in the same receiver stages; the internal noise of these stages is considered by supplying (converting) it to amplifier input and including it within additive noise $u_{\text{ш}}(t)$ within input signal-plus-noise y(t).

A second reason disrupting the validity of assumption 3 may be the presence of feedback from receiver output to its input. Such feedback may occur, for instance,

if the designed receiver is included in a closed radio control system (in a missile target seeker system, for example). Presence of this feedback will lead to the fact that distribution $P(x, y) = P(x)P_{x}(y)$ turns out to be dependent on receiver output voltage $\gamma(t)$ and, consequently, also on needed decision rule $\gamma = \Gamma(y)$. Therefore, optimum decision rule $\Gamma_{кр}(y)$ found without considering feedback may already not be optimum when this feedback is considered.

Strict mathematical system design of optimum receivers included in a closed control system turns out to be significantly more complex (see [133], for example]. However, in a number of cases, results obtained through closed system design turns out in the first approximation to be valid also when the receiver operates as part of a closed control system (see [144], for example).

A third reason for disruption of assumption 3 validity arises during receiver system design given <u>organized</u> (intentional) noise action.

An enemy will strive to create interference that would act the strongest against a system we have designed. Therefore, if we succeed in finding some algorithm $\Gamma_{кр}(y)$, which minimizes interference of a <u>given</u> type, then the enemy will attempt to change the nature of the interference in order to intensify its action. Therefore, in this case, called a conflict (adversary) situation, optimum receiver algorithm $\Gamma_{кр}(y)$ will depend input signal-plus-noise y(t) characteristics, while these characteristics in turn are dependent on the receiver action algorithm.

Receiver system design in conflict situations is complicated greatly. The most-typical system design approaches in such situations are:

1. Use of the approximation by iteration method, which consists of the following stages.

In the first stage, noise characteristics are supplied based on a priori /378
data and "common sense" and the optimum receiver structure corresponding to these characteristics is found, i. e., algorithm $\Gamma_{0\,кр}(y)$, which may be considered the zero approximation.

In the second stage, the noise (within the boundaries of given rational

limitations) which will exert the strongest interference for resultant structure $\Gamma_{0\,\text{кр}}(y)$ and the following (first) approximation $\Gamma_{1\text{кр}}(y)$ for optimum receiver structure is found considering this noise. Next, where required, second approximation $\Gamma_{2\text{кр}}(y)$ and so on may be found analogously. However, one should note that, in a number of cases, approximation by interation convergence turns out to be too slow or completely is lacking.

2. Refusal to perform mathematical system design, performance of a comparative analysis of the noise immunity of different specific receiver construction variants (given several organized noise characteristic variants), and selection of the most-suitable variant.

3. Use of game theory. Possible problem formulations when game theory is used are examined in the following section.

20.4 Problem Formulation When Game Theory is Used

Game theory is most advanced for so-called zero-sum pair games. Here, the receiver system design problem being examined is formulated in the following manner. Only two players participate in the game (pair game)--receiver designer (H_1) and organizer of the interference disrupting the operation of the receiver (H_2). The interests of the players are directly contradictory: a win for one of them is a loss for the other (zero-sum game). Each player will strive to play in such a way as to insure the best result for himself in the supposition that the opponent will strive to make this best result the worst. Consequently, game theory provides an opportunity to find the best algorithm for the worst potentially-possible case.

Games are divided into discrete (matrix) and analog (differential). In turn, discrete games are divided into single-step and multistep. In a single-step (single-move) game, the opponents make a total of one move each and the game ends. As opposed to this, a multistep (multimove) game comprises a sequence of moves made by each opponent. The theory of multistep games is considerably more complex and closely approximates the theory of dynamic programming.

In the case of a discrete single-step game, the task of the first player

boils down to selection of some strategy (move) γ_ν from a finite number M of possible moves, while the task of the second player is to select some strategy (move) β_μ of N possible moves (ν = 1, 2, . . ., M; = 1, 2, . . ., N). The results of each possible combination of moves (ν, μ) are evaluated by some /379 number $a_{\nu,\mu}$, equating to a win for the first player, and thereby a loss for the second. The following win (loss) matrix corresponds to the aggregates of all possible combinations of opposing strategies

$$A=\|a_{\nu,\mu}\|,$$

called a payoff matrix.

For example, this situation may be reduced to a discrete single-step game. M possible receiver action algorithms $\gamma_1 = \Gamma_1(y), \ldots, \gamma_\nu = \Gamma_\nu(y), \ldots, \gamma_M = \Gamma_M(y)$ exist. It is known that the enemy may use any of N possible organized interferences $\beta_1, \ldots, \beta_\mu, \ldots, \beta_N$ against the receiver. Since all possible algorithms Γ_ν and all possible interferences β_μ are assumed to be known beforehand, it is possible through analyis to estimate result $a_{\nu,\mu}$ of the action of interference β_μ on receiver with algorithm Γ_ν . The corresponding receiver quality index may be selected here as $a_{\nu,\mu}$.

For example, if the receiver is designed for signal detection, composite signal detection error probability P_{om} (or 1 -- P_{om}) may be such an index. The problem is to select from all possible strategies (algorithms) $\Gamma_1, \ldots, \Gamma_M$ the one which provides the best (the least if a = P_{om} or the greatest if a = 1 -- P_{om}) value of quality index $a_{\nu,\mu}$, given that the enemy uses the most dangerous (in the sense of the same quality index) of interferences $\beta_1, \ldots, \beta_N$.

For specificity, we will assume that quality index $a_{\nu,\mu}$ is such that the less it is, the better the receiver (i. e., the greater the win for the first player). For instance,

$$a_{\nu,\mu}=R_{\nu,\mu},$$

where $R_{\nu,\mu}$ -- average risk determined given that type β_μ interference acts on

the receiver and algorithm Γ_v is selected. Then, the problem may be formulated mathematically in the following manner.

Average risk $R_{v,\mu}$ magnitude will depend on selected receiver algorithm Γ_v and interference type β_μ, i. e.,

$$R_{v,\mu}=R(\Gamma_v,\ \beta_\mu). \tag{20.1}$$

Since fully-determinate v and μ correspond unambiguously to numbers Γ_v and β_μ, instead of (20.1), it also is possible to write

$$R_{v,\mu}=R(v,\ \mu). \tag{20.1a}$$

The first player will strive with Γ_v selection to minimize magnitude $R_{v,\mu}$ for such noise β_μ at which this minimal magnitude is maximum. For every strategy Γ_v, the worst (most dangerous) strategy β_v will be that for which

$$R(\Gamma_v,\ \beta_\mu)=\max_{\beta_\mu} R(\Gamma_v,\ \beta_\mu).$$

Therefore, the first player must select strategy Γ_v so that magnitude /380 $\max\limits_{\beta_\mu} R(\Gamma_v,\ \beta_\mu)$ will be minimal, i. e., from the condition that

$$\max_{\beta_\mu} R(\Gamma_v,\ \beta_\mu)=\min_{\Gamma_v}[\max_{\beta_\mu} R(\Gamma_v,\ \beta_\mu)]=R_1. \tag{20.2}$$

The second player, reasoning in an analogous manner, must select strategy β_μ from the condition

$$\min_{\Gamma_v} R(\Gamma_v,\ \beta_\mu)=\max_{\beta_\mu}\left[\min_{\Gamma_v} R(\Gamma_v,\ \beta_\mu)\right]=R_2. \tag{20.3}$$

In other words, the first player will start from the minimax criterion (minimizes maximum magnitude of risk R), while the second starts from the opposite, the maximin criterion (maximizes the minimum magnitude of risk R).

Since dependence $R(\Gamma_v, \beta_\mu)$ is assumed known, then it is possible from relationship (20.2) to find strategy β_μ' providing the risk R maximum and strategy Γ_v', which minimizes this maximum. Consequently, performing operations stipulated by relationship (20.2) makes it possible to find opposing strategy values β_μ' and Γ_v'.

Analogous to this, opposing strategy values β_μ'' and Γ_ν'' are obtained from relationship (20.3). Here, average risk value equalling R_1 corresponds to strategy combination $(\beta_\mu', \Gamma_\nu')$, while value R_2 corresponds to combination $(\beta_\mu'', \Gamma_\nu'')$. It is shown in game theory that always

$$R_1 \geqslant R_2. \tag{20.4}$$

If it turns out that

$$R_1 = R_2, \tag{20.5}$$

then, here

$$\beta_\mu'' = \beta_\mu' \quad \text{and} \quad \Gamma_\mu'' = \Gamma_\mu', \tag{20.6}$$

and, consequently, game theory makes it possible unequivocally to determine both the optimal strategies of both players, and the corresponding game outcome (game value), i. e., magnitude R.

However, it may happen when solving some problems that

$$R_1 > R_2. \tag{20.7}$$

In this case, the aforementioned approach does not make it possible to find either optimal strategies or game value and, obtaining a decision requires modification in the problem formulation, introducing in place of "pure" (determinate nonrandomized) strategies Γ_ν, β_μ so-called "mixed" (stochastic randomized) strategies. Here, players select strategy Γ_ν and β_μ randomly (through use of the corresponding independent statistical mechanisms) in accordance with probability distributions

$$\left.\begin{aligned} p &= (p_1, \ldots, p_\nu, \ldots, p_M), \\ q &= (q_1, \ldots, q_\mu, \ldots, q_N), \end{aligned}\right\} \tag{20.8}$$

where

$$\sum_{\nu=1}^{M} p_\nu = 1, \quad \sum_{\mu=1}^{N} q_\mu = 1.$$

When such mixed strategies are used, magnitude $R(\nu, \mu)$ turns out to be random and the game outcome is estimated by expected value

$$\overline{R} = \overline{R}(p, q) = \sum_{\nu=1}^{M} \sum_{\mu=1}^{N} R_{\nu,\mu} p_\nu q_\mu. \tag{20.9}$$

Here, player H_2 will strive to select distribution q so that magnitude $\overline{R}(p, q)$ will be maximum, while player H_1 selects distribution p so that this maximum magnitude will be the least. Therefore, in a case where mixed strategies are used, the following relationships are the analog of expressions (20.2) and (20.3), respectively:

$$\max_q \overline{R}(p, q) = \min_p \left[\max_q \overline{R}(p, q)\right] = R_1', \tag{20.2a}$$

$$\min_p \overline{R}(p, q) = \max_q \left[\min_p \overline{R}(p, q)\right] = R_2'. \tag{20.3a}$$

Here, the result always is

$$R_2' = R_1' = R', \tag{20.10}$$

where

$$R_1 \geqslant R' \geqslant R_2. \tag{20.11}$$

It follows from (20.10) that, when mixed strategies are used, a game (of the examined class) always has an outcome (i. e., fully-determinate optimum distribution p and q values and corresponding game outcome value R' exist), which may be found from relationship (20.2a) or (20.3a).

It follows from (20.11) that, when mixed (rather than pure) strategies are used, each of the players may not worsen his outcome and, where $R_2 > R_1$, he may

even improve it. Therefore, if use of pure strategies will lead to the inequality of magnitudes R_1 and R_2, the switch from pure to mixed strategies is completely logical.

Insurmountable difficulties usually are not encountered when finding outcomes of discrete games of the class examined. An outcome with the direct method is possible, given a small number of possible strategies (small M and N), while linear programming methods are used when these numbers are large. In complex cases, recommendations made in [132] are based on use of the corresponding pattern algorithms.

Analog (differential) games also may be based on use either of pure or of mixed strategies. But, in the case of analog games, the possible algorithm types $\gamma = \Gamma(y)$ are not limited to a finite number of discrete values, but may vary continuously (within the range of given limitations). This circumstance /382 increases mathematical difficulties in problem solution to such an extent that, at present, methods for solving analog games have been developed only for certain particular cases.

At present, a great deal of literature has been devoted to game theory [140, 141, 142, 143, and others]. However, its use for radio receiving system design still is very limited (see [171], for example). This mainly is explained by the following.

First, the mathematical "game" model described above only very roughly describes an actual adversary situation arising when providing electronic countermeasures.

Second, this mathematical model makes it possible to find the optimum receiver structure only in the supposition of the worst possible interference. If this interference in actuality does not turn out to be the worst, then the resultant receiver structure will not turn out to be optimum.

Third, solution of the problem in the case of an analog game may encounter serious mathematical difficulties.

However, given a conflict (adversary) situation, game theory, in spite of

its limitations and shortcomings, at present is the mathematical vehicle corresponding most closely to the nature of this situation.

20.5 Certain Additional Limitations Inherent in Statistical Decision Theory

We now will dwell on assumptions and limitations 4, 5, and 6 noted on pages 538 and 539.

Several limitations are placed on a receiver during its actual development (design process), the main ones usually being cost, weight, and overall dimensions. During receiver system design based on the aforementioned statistical decision theory, these very-important restrictions and quality indicators are not considered quantitatively at all. Only a certain hidden (qualitative) consideration of these limitations is possible. It may comprise, for example, the following:

1. Optimum algorithm $\Gamma_{\text{кр}}(y)$ is sought, not in the class of all physically realizable, but only in the class of determinate (nonrandomized) algorithms. This signifies that statistical (randomized) algorithms, usually (but not always) being more complex (and, consequently, having greater cost, weight, and overall dimensions) are not considered.

2. It is possible to seek optimum algorithm $\Gamma_{\text{кр}}(y)$ only in the class of linear systems (as is done in § 2.2, for example). Since linear algorithms usually (but not always) are simpler for realization, then this will lead to a decrease in cost, weight, and overall dimensions.

3. It is possible to seek the optimum algorithm in the single-channel, rather than multichannel, system class.

However, first, not in all cases is a nonrandomized algorithm simpler than a randomized one, a linear system simpler than a nonlinear one, or a single- /383 channel system simpler than a multichannel one. Second, there are other ways to simplify the system. Finally, there usually is a requirement to estimate the possible decrease in cost, weight, and overall dimensions both qualitatively and quantitatively.

For these reasons, optimum algorithm $\Gamma_{\text{кр}}(y)$ found from statistical decision

theory usually may not be considered final and requires further estimation and adjustment considering other quality indicators, primarily reliability, cost, weight, and overall dimensions.

A statistical decision theory (in its contemporary form) limitation is not only that it does not consider any receiver qualities other than message fidelity at all, but also that the latter is considered only by a single number--average risk R.

By definition, average risk equals the expected value of loss $l(x, \gamma)$ arising when decision γ is taken about message x. But, the expected value is far from the complete characteristic of random magnitude (or random process) $l(x, \gamma)$. In particular, besides expected value $R = \bar{l}$ of the loss magnitude, also of significant value may be this magnitude's variance

$$\sigma_l^2 = \overline{(l - \bar{l})^2}.$$

Therefore, a population of two quality indices--average risk R and risk variance σ_l^2, is a more-complete (but, in turn, <u>not</u> absolutely complete) message fidelity characteristic. However, optimum receiver system design based on a population of two or more indices is in the general case considerably more complex. Several special features of this system design are examined in Chapter 22.

To conclude this chapter, we will note the following. Statistical decision theory makes it possible to find structure $\Gamma_{\text{кр}}(y)$ of a receiver providing minimum average risk value $R = R_{\text{мин}}$. When seeking this structure, no consideration is given to a series of very-important factors [a priori data ambiguity, nonideality of the algorithm $\Gamma_{\text{кр}}(y)$ realization, presence of an entire series of additional quality indices such as cost, weight, overall dimensions, reliability, loss variance σ_l^2, and others]. However, consideration of any of these factors (or the entire aggregate of factors) may not lead to a decrease in average risk R compared to $R_{\text{мин}}$; it may cause only an <u>increase</u> in average risk. Therefore, one may assert that statistical decision theory makes it possible to find the potential (best theoretically-possible) value of the average risk magnitude which may not be exceeded (improved) in any real system, considering all its inherent limitations and quality indices.

Knowledge of this potential main receiver quality index value is very important during design of actual radio receiving devices. Therefore, statistical decision theory, in spite of all its serious inherent limitations, has found very wide use in radio receiving device system design.

CHAPTER TWENTY-ONE /384

BRIEF DESCRIPTION OF RECEIVER SYSTEM DESIGN METHODS GIVEN INCOMPLETE A PRIORI INFORMATION

21.1 General Description of Possible System Design Problem Formulation Variants and Ways to Solve Them

Use of "classic" statistical decision theory presented in Chapter 17 requires knowledge of the a priori distribution of message P(x) and likelihood function $P_x(y)$. The latter, in turn, is determined from the distributions of parasitic signal parameters (non-additive noise) and additive noise [see relationships (17.2)--(17.6)]. If all these distributions are known, then a so-called case of complete a priori information occurs and the optimum receiver structure may be found by average risk R minimization, i. e., from condition (17.29).

Essentially, however, all or some of these distributions usually are not known precisely*, i. e., a so-called case of incomplete a priori information occurs. In other words, presence of complete a priori information may be considered as a threshold (extreme) particular case of a real situation in which a priori information is always incomplete to some degree.

*In addition, in several cases, a priori distributions may in the process of receiving device operation change in accordance with unknown a priori laws.

Since phenomena in nature vary infinitely, the degree of a priori information incompleteness also may vary within unlimited bounds. Therefore, it is impossible in principle to create a single theory which would encompass all possible cases of a priori information incompleteness and thus be efficient enough in all cases. One may only indicate individual classes of phenomena and some theoretical approaches corresponding in some degree or other to each of these classes or some group of classes. Therefore, at present, the theory of optimum reception methods, given incomplete a priori information, still is in the initial development stage, in spite of the fact that, in recent years, many works have been devoted to it [131, 132, 137, 150, 151, 160-170, and others].

This chapter has as its goal only to familiarize the reader with the ideas and methods which make it possible, to a certain degree, to surmount this so-called a priori difficulty. Several ways of overcoming difficulties arising when a priori distribution P(x) of messages is unknown already have been pointed out in preceding chapters.

Thus, it was demonstrated in Chapter 19 that, given an increase in message fidelity (validity), the influence of distribution P(x) type on optimum receiver structure and properties will strive towards zero if the signal is precisely known or has safe parasitic parameters. Therefore, given sufficiently-high message /385 fidelity (reliability), it is possible in some cases to assume (in the first approximation at least) that distribution P(x) is uniform. If, along with distribution P(x), loss function $l(x, \gamma)$ also is unknown, then it is permissible in some cases, as noted above, to use the maximum likelihood criterion rather than the average risk criterion.

Another possible way is use of the minimax criterion, i. e., determination of the least-favorable a priori distribution $P_M(x)$ in which minimum average risk is maximum. However, this approach also is not always advisible because the actual distribution may be significantly more favorable also due to mathematical difficulties that arise.

In binary reception, when the message has only two values, x_0, and x_1, in a case of unknown probabilities $P(x_0)$ and $P(x_1) = 1 - P(x_0)$, it is possible to avoid characterizing reception quality by a single index—average risk R, and

to characterize it by two indices—conditional risks R_{x_0} and R_{x_1}. In a case of binary signal detection and unitary detection error weights, the result is $R_{x_0} = P_{\text{л т}}$ and $R_{x_1} = P_{\text{пр}}$, and the problem boils down to finding the structure of a receiver providing $P_{\text{пр}}$ = min where $P_{\text{л т}}$ = const (Neyman-Pearson criterion).

All these approaches are based on complete disregard for the a priori distribution P(x) type. However, in several cases, especially in filtration, this complete disregard may be risky, even given relatively-high message fidelity. Here, in several cases, the following method may turn out to be useful.

Let the requirement be to extract message x(t) from additive mixture

$$y(t) = x(t) + u_{\text{ш}}(t). \qquad (21.1)$$

It was shown in § 2.2 that, when the optimum system is sought in the class of linear systems and the loss function selected is quadratic, optimum system structure and properties completely are determined by correlation functions $R_x(t_1, t_2)$ and $R_{\text{ш}}(t_1, t_2)$*. Consequently, given this system design problem formulation, there is no requirement to know complete multidimensional distribution P(x) of random process x(t). It suffices to know its correlation function $R_x(t_1, t_2)$. Therefore, if mixture y(t) has the (2.1) form and nothing is known about message x(t), with the exception of $R_x(t_1, t_2)$, it is advisible to use the same designed system class limitations and loss function.

Up to this point, we have examined only that part of the a priori difficulty linked with complete or partial a priori message distribution uncertainty. However, in the general case, parasitic signal parameter and additive noise distributions may be unknown to a certain degree as well. Therefore, in more-general form, the problem may be formulated as follows.

In a case of complete a priori information, distribution /386

$$P(x, y) = P(x)P_x(y),$$

*If processes x(t) and $u_{\text{ш}}(t)$ are interrelated, then one also must know $R_{x\text{ш}}(t_1, t_2)$.

required to compute average risk R is precisely known. In a case of incomplete a priori information, distribution P(x) may be known, with the exception of the parameter population

$$\vec{\nu} = \{\nu, \ldots, \nu_m\}, \tag{21.2}$$

while distribution $P_x(y)$ is known, except for parameter population

$$\vec{\mu} = \{\mu_1, \ldots, \mu_l\}. \tag{21.3}$$

This means that the only knowns are conditional distributions

$$P_{\vec{\nu},\vec{\mu}}(x, y) = P_{\vec{\nu}}(x) P_{\vec{\mu},x}(y). \tag{21.4}$$

The degree of parameter $\vec{\nu}$ and $\vec{\mu}$ uncertainty may differ. Here, the following cases may occur:

1. Unknown parameters $\vec{\nu}$ and $\vec{\mu}$ are random magnitudes (or processes), the laws of distribution of which $P(\vec{\nu})$ and $P(\vec{\mu})$ are precisely known.*

2. Unknown parameters $\vec{\nu}$ and $\vec{\mu}$ are random magnitudes (or processes), i. e., distributions $P(\vec{\nu})$ and $P(\vec{\mu})$ exist; however, as opposed to the first case, these distributions are unknown.

3. Parameters $\vec{\nu}$ and $\vec{\mu}$ a fortiori are non-random (determinate) unknown magnitudes (or time functions).

4. It also is unknown whether parameters $\vec{\nu}$ and $\vec{\mu}$ are random or not.

In all these cases, the assumption is that the type of laws of distribution P(x) and $P_x(y)$ is known; the only unknowns are some parameters $\vec{\nu}$ and $\vec{\mu}$ of these laws (expected value and variance, for example). However, there are cases

---

*If there is a statistical link between $\vec{\nu}$ and $\vec{\mu}$, then their joint distribution $P(\vec{\nu}, \vec{\mu})$ is known.

when even the law of distribution types is unknown. These extreme (of all those examined above) cases, in the sense of a priori information incompleteness, in mathematical statistics usually are called nonparametric. The following problem of binary detection of a signal on a noise background may serve as an example of a nonparametric problem.

The requirement is to detect whether or not a signal is present from analysis of sample $y = (y_1, \ldots, y_i, \ldots, y_n)$ comprising n independent sample values y_i. Nothing is known about law of distribution (integral) $F(y_i)$, except that this distribution is even when there is no signal and parity is disrupted /387 when there is a signal. As will be demonstrated in § 21.3, given a large sample, n, even such minimal a priori information may suffice for signal detection.

Especially complex are cases when it is unknown whether distributions are stationary (i. e., constant over time) or non-stationary or when it is known that they change over time, but the law of this change a priori is unknown. In future, we will assume that unknown distributions may be considered stationary (or changing over time in accordance with a priori known laws).

Finally, special note should be taken of a case of intentional (organized) interference. In this case, during receiver system design, it is known that the enemy will premeditatively change interference distributions, but it is unknown just how this will be accomplished. In such an adversary situation, as noted in the previous chapter, game theory is the most-adequate mathematical vehicle. Since problem formulation in game theory already is examined above, in this chapter we will examine only unintentional interference.

Preliminary (prior to receiver system design) receipt of unavailable a priori information through accumulation and processing of the corresponding experimental statistical material is the most-evident and -radical way to surmount an a priori difficulty. If, in accordance with problem conditions, it is possible to take these measures, it is advisible in a majority of cases to do so since an increase in a priori information about messages, signals, and noise makes it possible to improve message fidelity. However, in a number of cases, preliminary supplementation of unavailable a priori information is impossible in principle or due to a shortage of time and difficulty in performing experiments.

Another approach lies in creation of possibilities for preliminary patterning of the receiving device, using "demonstration" methods. Here, during the pattern process, not only some realization $y_0(t)$ is supplied to receiver input, but, simultaneously, it is indicated (noted) whether there is a signal in this realization or not (in detection problems), and, if so, then exactly which signal (in signal discrimination problems). In the general case, it is possible to conduct a demonstration, not of single realization $y_0(t)$, but a population of several realizations,

$$\overrightarrow{y_0(t)} = \{y_0^{(1)}(t), \ldots, y_0^{(m)}(t)\},$$

with simultaneous indication of the special features of each realization (for example, indication that there is absolutely no signal in realization $y_0^{(1)}(t)$, signal $u_{c1}(t)$ will be contained in realization $y_0^{(2)}(t)$, signal $u_{c2}(t)$ will be contained in realization $y_0^{(3)}(t)$, only the signal carrier oscillation will be contained in realization $y_0^{(4)}(t)$, i. e., there is no modulation by the measured message, and so forth).

Consequently, during patterning, the decision on message x reproduction will be taken in the general case based on two realization populations, $\overrightarrow{y_0(t)}$ /388 and $\overrightarrow{y(t)}$, where $\overrightarrow{y_0(t)}$ -- population of pattern realizations, $\overrightarrow{y(t)}$ -- population of operating realizations. Operating population $\overrightarrow{y(t)}$ reaches receiver input as the receiver operates under actual conditions rather than pattern conditions. Therefore, it may not be accompanied by pattern indices.

Evidently, the pattern process will be effective only when there is confidence that noise, signal, and message statistical characteristics remain unchanged during transition time from the pattern to the actual operating process. Therefore, in a number of problems, it is possible to require realizations $\overrightarrow{y_0(t)}$ and $\overrightarrow{y(t)}$ to be separated over time by relatively-slight intervals or even be transmitted sequentially in parts (or simultaneously if proper division by frequency or by another feature is possible).

In radio communications systems, the patterning idea began to be realized many decades ago by sending so-called pilot signals, i. e., an oscillation not modulated by the transmitted message and close in frequency to the operating signal carrier frequency. Based on distortions of the pilot signal reaching receiver

input, it is possible to judge the nature and intensity of non-additive (modulated) noise arising in the communications channel.

Since that time, various pattern principles to a certain degree are being used in other fields of radio electronics as well. However, if until recently pattern algorithm selection has been intuitive, at present, statistical decision theory, the theory of stochastic approximation and other methods of mathematical statistics began to be used in their design. Examples of such an approach are given in § 21.4 and 21.5.

However, in a number of cases, patterning is impossible or inadvisible, due to equipment complexity arising at this time, for instance. In such cases, the decision about the reproduced message must be taken only on the basis of operating realization y(t) [or $\overrightarrow{y(t)}$] and a priori information incompleteness impacts to the greatest degree. It is all the more important under such difficult conditions to find, where possible, the optimum receiver action algorithm. However, solution of this problem, as already noted, will be found in the initial stage and results obtained mainly for a case of binary signal detection.

At present, the following basic optimum receiver system design methods are known for incomplete a priori information (excluding game theory, already examined in Chapter 20):

1. Methods which are a development of statistical decision theory.

2. Nonparametric methods of mathematical statistics.

3. Methods based on the theory of stochastic approximation.

A brief description of these methods is provided in the following sections.

21.2 Methods Developed From Statistical Decision Theory /389

1. General Description

Methods developed from statistical decision theory are used in cases when

a priori distribution $P(x, y) = P(x)P_x(y)$ is known (or assumed known), with the exception of some parameter population ($\vec{\nu}, \vec{\mu}$) [see (21.2)-- (21.4)], i. e., the only known is a conditional distribution of the (21.4) type, which may be designated $P_{\vec{\theta}}(x, y)$, where $\vec{\theta} = \{\vec{\nu}, \vec{\mu}\}$ -- population of all unknown distribution parameters.

If a priori distribution $P(\vec{\theta})$ of vector $\vec{\theta}$ is known, then it is possible to find unconditional distribution P(x, y) from the formula

$$P(x, y) = \int_{A_{\vec{\theta}}} P_{\vec{\theta}}(x, y) P(\vec{\theta}) d\vec{\theta} \tag{21.5}$$

(where $A_{\vec{\theta}}$ -- range of all possible vector $\vec{\theta}$ values) and to use it in the normal manner to compute average risk R.

For example, relationship (17.29) is valid for determinate (non-randomized) decision rule (17.7) and, considering (21.5), optimum decision rule $\Gamma_{\text{кр}}(y)$ may be found from the condition

$$R = \int_{A_x} \int_{A_y} \int_{A_{\vec{\theta}}} l(x, \Gamma(y)) P_{\vec{\theta}}(x, y) P(\vec{\theta}) dx\, dy\, d\vec{\theta} = \min. \tag{21.6}$$

In the general case, when decision rule $\Delta(\gamma/y)$ may be randomized, instead of (21.6), one should assume that

$$R = \int_{A_x} \int_{A_y} \int_{A_\gamma} \int_{A_{\vec{\theta}}} l(x, \gamma) P_{\vec{\theta}}(x, y) P(\vec{\theta}) \Delta(\gamma | y) dx\, dy\, d\gamma\, d\vec{\theta} = \min. \tag{21.7}$$

Hence, from the point of view of principle, it follows that statistical decision theory easily is generalized for a case when population $\vec{\theta}$ of the parameters of a priori distribution P(x, y) is unknown, but does exist, and distribution $P(\vec{\theta})$ of these parameters is known. However, here, computational difficulties may rise sharply here. In addition, in many cases, distribution $P(\vec{\theta})$ is unknown; therefore, we will switch to examination of just such cases.

One possible way to surmount a priori difficulties is to use the minimax method, which boils down to the following in this case.

A determination is made of the least-favorable type $P_M(\vec{\theta})$ of this distribution, i. e., that type of function $P(\vec{\theta})$ in which the minimum average risk R value [determined from relationship (21.6) or (21.7)] is maximum. That decision /390 rule $\Gamma(y)$ [or $\Delta(\dot{\gamma}/y)$] which provides minimum average risk R for this distribution $P_M(\vec{\theta})$ is optimum. However, mathematical difficulties in determining (computing) distribution $P(\vec{\theta})$ in a number of cases essentially are insurmountable. In addition, as repeatedly noted, the minimax criterion sometimes is too "dangerous" since the actual distribution in many cases may differ considerably from the least favorable one. Therefore, besides the minimax method, other methods for overcoming an a priori difficulty also are of interest.

The simplest method is simply to assume the unknown distribution in expression (21.6) or (21.7) for average risk is uniform, i. e., to consider that

$$P(\vec{\theta}) = \text{const.} \tag{21.8}$$

However, in a number of cases, especially in cases of filtration, this may lead to impermissible system design errors. Therefore, several authors [160, 162, and others] propose that a different tack be taken, this being to require that a designed receiver provide an estimate of all unknown parameters $\vec{\theta}$ along with message x reproduction. Then, with respect to problem conditions, if it turns out that the estimate of these parameters is very accurate, there is a further basis on which to consider that a priori distribution $P(\vec{\theta})$ type impacts little on reception results and, consequently, may be considered uniform. This assertion to a significant degree is analogous to the Chapter 19 supposition that the higher the message x fidelity, the less distribution P(x) type impact and all the more reason to consider this distribution uniform.

It is possible to explain the utility of introducing estimation of unknown parameters $\vec{\theta}$ also in the following manner. When such an estimate is given, a fortiori distribution $P_y(\vec{\theta})$ is computed directly or indirectly in the receiving device and it is considered when decision γ is taken regarding reproduced usable message x. This "smooths" the arbitrariness that a priori distribution $P(\vec{\theta})$ is assumed to be uniform. Evidently, if the ideally-precise estimate of parameters $\vec{\theta}$ is obtained, then the "smoothing" also would be ideal, i. e., the assumption concerning the uniformity of a priori distribution $P(\vec{\theta})$ would not introduce any additional error into message x reproduction.

However, it still does not follow from this that replacement of the simple assumption about distribution P($\vec{\theta}$) uniformity by an estimate of parameters $\vec{\theta}$ always is advisible. Actually, it follows from the material presented in preceding chapters, for example, that, when measuring signal amplitude (or during signal detection), the law of distribution of constant (but random) signal initial /391 phase φ_0 does not influence reception quality if it is high (i. e., signal-to-noise ratio is high). Therefore, during design of a receiver system operating under such conditions, if the requirement is simultaneous (with detection) initial phase φ_0 estimate, then this will lead only to its unjustified complexity.

But, conversely, it is evident that, in a number of cases, introduction of an additional estimate of unknown signal parameters or noise may provide a radical increase in reception quality. For example, when measuring the amplitude of a signal with a carrier frequency unknown within broad limits but stable, much higher amplitude measurement accuracy may be obtained in the steady-state mode if the receiver performs, in addition, signal carrier frequency measurement and, consequently, a radical narrowing of its bandwidth is possible. Therefore, the method of overcoming an a priori difficulty through introduction of an additional estimate of parasitic signal parameters may be useful in a number of cases. From the mathematical point of view, problem formulation will boil down to the following.

We assume that the designed receiver must perform not only estimate (reproduction) γ of usable signal x, but estimate $\vec{\theta}^*$ of unknown parameters $\vec{\theta}$ of distribution $P_{\vec{\theta}}(x, y)$. For simplicity, let the estimate algorithms be determinate (non-randomized), i. e.,

$$\gamma = \Gamma(y) \quad \text{and} \quad \vec{\theta}^* = B(y). \tag{21.9}$$

Estimate quality is characterized by generalized loss function $l(x, \vec{\theta}, \gamma, \vec{\theta}^*)$, which denotes losses with varied combinations of true values x and $\vec{\theta}$ and of their estimates γ and $\vec{\theta}^*$. Then, analogous to (21.6), it is possible to write the following receiver optimization:

$$R = \int_{A_x}\int_{A_y}\int_{A_{\vec{\theta}}} l(x, \vec{\theta}, \Gamma(y), B(y)) \times \times P_{\vec{\theta}}(x, y) P(\vec{\theta})\, dx\, dy\, d\vec{\theta} = \min. \tag{21.10}$$

Here, the minimum is sought with respect to all possible values of operators $\Gamma(y)$ and B(y).

Finding, from this condition, optimum values of algorithm (decision rules) $\Gamma(y)$ and B(y) requires knowing a priori distribution $P(\vec{\theta})$, which is unknown to us. Therefore, it is assumed uniform:

$$P(\vec{\theta}) = \mathrm{const}$$

(qualitative substantiation of this assumption was provided above). Here, mathematically, the problem solution is completely determinate and, consequently, possible in principle. However, great mathematical difficulties and a requirement to /392 substantiate the complex loss function $l(x, \vec{\theta}, \gamma, \vec{\theta}^*)$ type remain.

Therefore, in one work [162], an additional assumption is made for the sake of simplicity: the assumption is that, given high parameter x and $\vec{\theta}$ estimate accuracy and a symmetrical loss function form, the type of this function will not play a role and one may assume that minimum average risk R [expression (21.10)] coincides with the maximum with respect to ($\vec{\theta}$, x) of distribution $P_{\vec{\theta}}(x, y)$. Here, optimum estimate algorithm values may be found from the following, simpler, condition:

$$\gamma = x_{\text{нр}}(y), \quad \vec{\theta}^* = \vec{\theta}_{\text{нр}}(y), \tag{21.11}$$

where $x_{\text{нр}}(y)$ and $\vec{\theta}_{\text{нр}}(y)$ -- values of parameters x and $\vec{\theta}$, for which probability density $P_{\vec{\theta}}(x, y)$, looked upon as a function of x and $\vec{\theta}$, has a maximum.

Evidently, this method is a generalization of the method presented in Chapter 17 based on replacement of the minimum average risk criterion by the maximum a priori (inverse) probability density criterion. However, in order to bring problem solution to a conclusion, even with respect to the simplified criterion (21.11), a series of additional assumptions usually must be made (see [162], for example).

The basic drawback of the aforementioned methods is the requirement for often insufficiently-substantiated replacement of unknown distribution $P(\vec{\theta})$ by a uniform

distribution. Therefore, the following group of methods in which this assumption is not required is of interest.

The group of methods is based on introduction into the requirements levied on the designed receiver of the additional requirement for slight dependence or complete independence of its structure and/or its properties on unknown distribution $P(\vec{\theta})$. Naturally, introduction of such a requirement (limitation), just like any additional requirement, in the general case worsens the achievable value of the basic receiver quality index (it increases average risk R, for example). However, there is no arbitrariness here linked with selection of the type of unknown distribution $P(\vec{\theta})$ and the corresponding arbitrariness in the actual basic quality index value.

This group of methods at present is more developed relative to binary signal detection and, depending on the type of additional requirement levied, is divided into the following sub-groups:

1. Methods based on the requirement for invariance.

2. Methods based on the requirement for similarity.

3. Methods based on the false-alarm probability upper bound.

Since false-alarm probability $P_{\text{лт}}$ and miss probability $P_{\text{пр}}$ are the basic receiver quality indices given binary detection and incomplete a priori information, the aforementioned requirements are formulated mathematically in the following /393 manner:

1. Invariance. Probabilities $P_{\text{лт}}$ and $P_{\text{пр}}$ must not depend on distribution $P(\vec{\theta})$ (or on $\vec{\theta}$).

2. Similarity requirement. Probability $P_{\text{лт}}$ must not depend on distribution $P(\vec{\theta})$ (or on $\vec{\theta}$), i. e., given a change in $P(\vec{\theta})$ (or $\vec{\theta}$), this condition must be met:

$$P_{\text{лт}} = \text{const} = P_{\text{лт}0}, \tag{21.12}$$

where $P_{\text{лт}0}$ -- given (permissible) false-alarm probability magnitude.

3. Requirement for a magnitude $P_{\text{лт}}$ upper bound. Given changes in $P(\vec{\theta})$ (or $\vec{\theta}$), this condition must be met:

$$P_{\text{лт}} \leqslant P_{\text{лт}0}. \tag{21.13}$$

It should be noted that, when any of the aforementioned conditions are met, so-called impermissible decision rules must be excluded from the examination, i. e., those algorithm $\gamma = \Gamma^*(y)$ types in which decision turns out independent of input realization y. [For example, it is simple to satisfy the invariance requirement if you simply shut off the receiver*; here, $P_{\text{лт}} = 0$ and $P_{\text{пр}} = 1$ will always be the case (for any $\vec{\theta}$). It is evident that such a decision rule is absolutely unsatisfactory].

Of the three requirement types indicated, the first is the most rigid and often is not satisfied at all (if you rule out impermissible decision rules), or will lead to a significant deterioration in quality indices $P_{\text{лт}}$ and $P_{\text{пр}}$. Requirement (21.13) is the least stringent and, in a number of cases, may be satisfied with less quality index deterioration.

In some particular cases (especially given a large independent observation sample size n), it turns out that invariance algorithms, similar algorithms, and algorithms based on an upper magnitude $P_{\text{лт}}$ bound coincide among themselves, as well as with optimum algorithms found in the conventional way for known distribution $P(\vec{\theta})$. It is evident that this occurs in those cases when independent a priori distribution $P(\vec{\theta})$ type does not impact on optimum receiver structure and properties. However, in the general case, results obtained using various methods do not coincide.

*Or there is no reception whatsoever or you assert that a signal always is absent.

Several results V. A. Korado obtained based on the similarity requirement will be presented below as an illustration [163--167].

2. An Example of System Design of Optimum Detectors Satisfying the Similarity Requirement /394

Consider binary detection of a signal on a background of gaussian noise. The assumption is that intensities ν and μ of signal and noise are the only unknown parameters characterizing distributions $P_{xo}(y)$ and $P_{x1}(y)$. Also unknown is the ratio of these intensities

$$q=\frac{\nu}{\mu}. \tag{21.14}$$

Not only magnitudes ν and μ are unknown, the laws of their distribution also are unknown; the only known is that these intensities are constant during detection time.

When a system is designed under these conditions, a limitation is imposed on the range of the possible signal-to-noise ratio values:

$$q\geqslant q_0, \tag{21.15}$$

where q_0 -- unknown <u>positive</u> magnitude. The requirement for this magnitude q limitation below some non-zero value q_0 is evident since, where $q = 0$, the signal a fortiori is absent and the signal detection problem loses its meaning.

The first requirement levied on the designed system is similarity requirement (21.12). The second requirement concerns miss probability P_{np}.

Given unknown ν and μ, just as in the case of known values for these parameters, miss probability P_{np} will depend, not on absolute ν and μ magnitudes, but only on their ratio q: the greater the q, the less the P_{np}. Here, it is desirable to find that decision rule $\Gamma_{opt}(y)$ which would provide the probability P_{np} minimum for <u>all</u> values $q \in A_q$ [where A_q -- range of all possible parameter q values determined from inequality (21.15)]. However, in the general case, such

a rule does not exist. Therefore, a more-modest requirement must be satisfied, namely a requirement of the type

$$\underset{q \in A_q}{\text{макс}} P_{\text{пр}}(q, \Gamma'_{\text{опт}}) \leqslant \underset{q \in A_q}{\text{макс}} P_{\text{пр}}(q, \Gamma). \tag{21.16}$$

Here, needed optimum decision rule $\Gamma'_{\text{опт}}(y)$ provides the minimum [with respect to all possible decision rules $\Gamma(y)$] of the miss probability maximum [with respect to all possible values of unknown parameter q]. This decision rule is the minimax rule and coincides with Bayes decision rule $\Gamma_{\text{нр}}(y)$ for least-favorable a priori distribution $P_M(q)$ [i. e., of distribution P(q) providing the maximum of the minimum value of average risk R determined from formula (21.6). Therefore, it suffices for finding rule $\Gamma'_{\text{опт}}(y)$ to find the Bayes decision rule for parameter distribution $P_M(q)$. Here, it turns out that the optimum decision rule has the following form:

decision "yes" (signal) will be taken if /395

$$l(y) = \frac{\int_{A_q} P_{x_1, q}(y) P_M(q)\, dq}{P_{x_0, \mu_0}(y)} > C(g_y); \tag{21.17}$$

decision "no" (no signal) conversely, i. e., when

$$l(y) \leqslant C(g_y).$$

Here, μ_0 -- random fixed value of noise intensity μ ; $g_y = g(y)$ -- sufficient [for estimation of parameter μ of distribution $P_{x_0, \mu}(y)$] statistic* (more precisely,

*In the theory of mathematical statistics (see [131], for example), function $g(y_1, \ldots, y_n)$ of sample value population $(y_1, \ldots, y_n)$, which satisfies the following factorization (special expansion) condition, is called a sufficient (for estimation of probability distribution parameter θ) statistic of random process y(t):

$$P_\theta(y_1, \ldots, y_n) = f[g(y_1, \ldots, y_n), \theta]\, h(y_1, \ldots, y_n), \tag{21.18}$$

where the second factor will not depend on θ . Using this relationship, it is possible to determine sufficient statistic g(y) corresponding to known conditional distribution $P_\theta(y_1, \ldots, y_n)$. For example, if $y_1, \ldots, y_n$ is independent and identically distributed in accordance with the Poisson law, then $P_0 = \dfrac{\theta^{\left(\sum_{i=1}^{n} y_i\right)} e^{-n\theta}}{\prod_{i=1}^{n} (y_i)!}$.

realization of the sufficient statistic) of input mixture y(t); $C(g_y)$ -- function of g_y found in accordance with similarity requirement (21.12) from the condition

$$P_{лт} = \text{const} = P_{лт0}$$

for all values of sufficient statistic g_y.

It is easy to become convinced from comparing the left side of inequality (21.17) with expression (14.44) that it is a likelihood factor averaged with respect to all possible unknown parameter q values. Function $C(g_y)$, standing on the right-

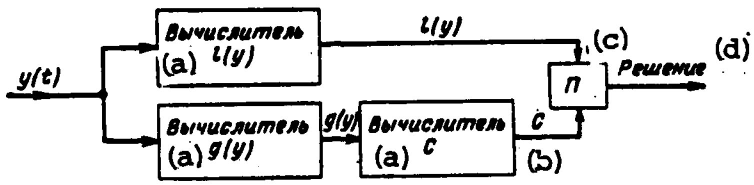


Figure 21.1. (a) -- Computer; (b) -- C [threshold comparator]; (c) -- P [threshold]; (d) -- Decision.

hand side of inequality (21.17), plays the role of some threshold with which likelihood factor l(y) value is compared. Therefore, the optimum detector functional diagram may be represented as shown in Figure 21.1. It differs from a conventional Neyman--Pearson detector only in that the likelihood factor is computed for least-favorable (rather than the actual) distribution $P_M(q)$, while threshold bias C is not constant, but is computed (with the aid of an additional channel) for every specific realization y(t).

If the unknown parameter (parameters) was not signal-to-noise ratio q /396 [or the population of intensities (ν, μ) but some other physical parameter (parameters) of mixture y(t), then computation of the least-favorable a priori distribution might present greater, sometimes even insurmountable, mathematical difficulties. However, in the case under examination, when this parameter is signal-to-noise ratio q (or the population of intensities ν and μ), distribution $P_M(q)$ may be found from the following evident considerations.

---

and, substituting this expression with the right-hand portion of relationship (21.18), we will find that $g(y) = \sum_{i=1}^{n} y_i$.

The less the q, the greater the miss probability P_{np}. Therefore, the least favorable is that distribution P(q) in which magnitude q always equals minimum-possible value q_o. Consequently,

$$P_{n}(q)=\delta(q-q_0), \tag{21.19}$$

where $\delta(q-q_0)$ -- δ-function.

Considering this relationship, algorithm (21.17) takes the form

$$\frac{P_{x_1, q_0}(y)}{P_{x_0, \mu_0}(y)} > C(g_y). \tag{21.20}$$

In many practical cases, it turns out that distribution $P_{x_0, \mu_0}(y)$ will depend on y only via sufficient statistic g_y. In this case, algorithm (21.20) is simplified even more and takes the form

$$P_{x_1, q_0}(y) > C'(g_y), \tag{21.21}$$

where $C'(g_y)$ also is some function of sufficient statistic g_y.

It should be noted that, in a majority of practical cases, there is no success in predetermined substantiation of signal-to-noise ratio q_o boundary selection. Therefore, it often is more proper to require, not minimization of the maximum miss probability P_{np} value occurring when $q = q_o$, but minimization of lower bound q_o corresponding to the given (permissible) magnitude $P_{np\,0}$ of miss probability P_{np}. This denotes that, instead of criterion (21.16), it is more logical to use the following criterion:

$$q_o = \min \text{ where } P_{np} \leqslant P_{np\,0} \text{ for all } q \geqslant q_o \tag{21.22}$$

[retaining here additional similarity requirement (21.12)]. Criterion (21.22) will lead to the identical (optimum) decision that criterion (21.16) does /397 [while retaining in both cases similarity requirement (21.12) with an identical probability $P_{\text{л т}}$ value] if the following correspondence exists between values q_o and P_{np} in both cases.

Using criterion (21.16), consider the case where, for given (known) value q_0, we found the minimum value of the maximum (i. e., occurring when $q = q_0$) miss probability $P_{\text{пр мин}}$ and corresponding optimum algorithm $\Gamma'_{\text{опт}}(y)$. Then, using criterion (21.22) for $P_{\text{пр}\,0} = P_{\text{пр мин}}$, we will obtain identical decision rule $\Gamma'_{\text{опт}}(y)$; minimum lower bound value $q_{0\,\text{мин}}$ corresponding to it will coincide with value q_0, which is given for criterion (21.16). Here, we have, in essence, the following example of a general system design problem based on two quality indices (this general problem is examined in the next chapter).

When using both criterion (21.16) and (21.22), there are two designed system quality indices which it is desirable to minimize: miss probability $P_{\text{пр}}$ and the value of signal-to-noise ratio q_0 lower bound. In the general (nondegenerative) system design case, it is impossible to select that decision rule $\Gamma(y)$ which would minimize $P_{\text{пр}}$ and q_0 simultaneously (see Chapter 22 also). Therefore, in the first case [criterion (21.16)], the problem posed will be to minimize magnitude $P_{\text{пр}}$ for a given (fixed) magnitude q_0 while, in the second case [criterion (21.22)], it is to minimize q_0 for a given value $P_{\text{пр}}$. In other words, in the first case, during the system design process, q_0 will be transferred to the category of limitations, while $P_{\text{пр}}$ is transferred in the second case. However, in the next chapter, it will be demonstrated that if, as a result of system design based on one of these methods, the dependence between both quality indices (in this case, $P_{\text{пр}}$ and q_0) will turn out to be strictly monotonic, then the system design result will not change if the other (second) quality index is transferred to the category of limitations instead of the given quality index.

In the case of indices q_0 and $P_{\text{пр}}$ we are examining, the dependence between them in the case of system design based on criterion (21.16) turns out to be strictly monotonic: the less the q_0, the greater the $P_{\text{пр}}$ and vice versa. Therefore, from the point of view of principle, criteria (21.16) and (21.22) turn out to be equivalent. However, from the practical point of view, the following differences exist between these two criteria: criterion (21.16) is more convenient from the point of view of the mathematical computations required to find optimum algorithm $\Gamma'_{\text{опт}}(y)$, while criterion (21.22) is more convenient from an engineering standpoint. Therefore, in similar cases, system design follows criterion (21.16) and will find the optimum link between quality indices $P_{\text{пр}}$ and q_0, i. e., a performance curve of the $P_{\text{пр}} = f_{\text{р}}(q_0)$ type. During computations, this curve is used, not to

determine magnitude P_{np} for a given q_0 (since magnitude q_0 usually a priori is unknown), but rather for determination of magnitude q_0 with which there is a guarantee of obtaining the maximum-permissible miss probability P_{np}, magnitude.

So, it is possible (given unknown intensities ν and μ) to use the optimization described by relationships (21.12) and (21.16) and leading to an optimum decision rule of the (21.20) or (21.21) type for synthesis of optimum similar /398 algorithms to detect and plot the optimum detector performance curve.

Based on these relationships, in [163 – 167] optimum detector systems were designed and their performance curves for several specific signal types were determined in the assumption that realization y(t) has the form of discrete sample ($y_1, \ldots, y_i, \ldots, y_n$). Here, as could be expected, by virtue of the increase is sample size n, power losses in the threshold signal (i. e., in magnitude q_0), obtained given unknown magnitude μ (compared to a case when magnitude μ is known), sharply decrease and, where $n \geqslant 25$—50, do not exceed 1 db. Conversely, when n is slight, losses may be very great, i. e., be several dozen decibels.

In the particular case when the signal has a normal law of distribution, the Figure 21.1 circuit takes the form depicted in Figure 21.2. In this circuit,

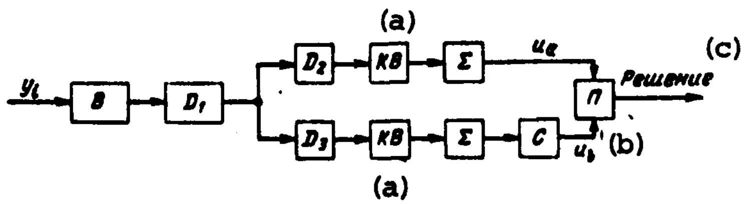


Figure 21.2. (a) -- KB [squarers]; (b) -- P [threshold comparator]; (c) -- Decision.

B -- linear filter (discrete); KB -- squarers; Σ -- adders of n input magnitudes; D_1, D_2, and D_3 -- linear systems with diagonal transformation matrices (in particular cases, all three or several of these units may be absent); C -- multiplier to constant number C; Π -- threshold comparator: if $u_a > u_b$, then the decision that there is a signal is taken, while when $u_a \leqslant u_b$, then the no signal decision

is taken. In both circuits (Figures 21.1 and 21.2), the presence of an additional (lower) channel will lead to an automatic threshold change (C or $u_{б}$) when noise intensity changes, which insures, in particular, that the requirement for false-alarm probability constancy is met [similarity requirement (21.12)].

21.3 Use of Nonparametric Methods of Mathematical Statistics

It already was noted in § 21.1 that nonparametric methods are used in those cases when not only distribution parameters, but its type as well, are unknown. Nonparameteric problems are complex because, given an unknown law of distribution, in the general case it is impossible in principle to compute the magnitude of the average risk or even of conditional risks. Consequently, it is impossible to estimate the quality inherent in given decision rule $\Gamma(y)$. For example, it is impossible in a case of binary signal detection to compute probabilities /399 $P_{лт}$ and $P_{пр}$. However, it turns out that, given certain problem formulations and appropriate decision rule selection, it is possible precisely to compute at least false-alarm probability $P_{лт}$, while one may only obtain a qualitative answer relative to miss probability $P_{пр}$. In particular, one usually is restricted to cases when sample $(y_1, \ldots, y_n)$ of n statistically-independent equally-distributed magnitudes, while sample size n is large enough, completely characterizes input realization y(t).

There is a significant number of nonparametric methods (see [131], for example]. We will examine only several of them, the ones which may be applied directly to binary signal detection problems.

The first is used in statistics to test the hypothesis of an unknown distribution parity and consists of the following.

Let sample $(y_1, \ldots, y_i, \ldots, y_n)$ of n independent and identically-distributed magnitudes be given. The requirement is to test hypothesis H that this sample will belong to distribution (integral) $F_1(y_i)$ relative to alternative K that the sample belongs to distribution $F_2(y_i)$. The types of functions $F_1(y_i)$ and $F_2(y_i)$ are unknown. The only known is that distribution $F_1(y_i)$ is even*,

*Positive and negative random magnitude values are equally probable.

i. e., it satisfies the condition

$$F_1(-y_i) = 1 - F_1(y_i), \tag{21.23}$$

while distribution $F_2(y_i)$ does not satisfy the parity condition, i. e.,

$$F_2(-y_i) \neq 1 - F_2(y_i). \tag{21.24}$$

The following procedure may be used to obtain acceptable decision rule $\Gamma(y)$. Analog random magnitude y_i is quantized into binary random magnitude u given a zero threshold, i. e., in accordance with the law

$$u(y_i) = \begin{cases} 1 & \text{where } y_i > 0 \\ 0 & \text{where } y_i \leqslant 0. \end{cases} \tag{21.25}$$

The sum of the sample values then is formed as follows

$$z = \sum_{i=1}^{n} u(y_i). \tag{21.26}$$

Since sample elements y_i are independent and equally distributed, then magnitude z is subordinate to a binomial law of distribution with parameters n and

$$p = 1 - F(0), \tag{21.27}$$

where $F(y_i)$ -- integral law of distribution.

If hypothesis H is valid, i. e., distribution $F(y_i)$ is even, then, in /400 accordance with (21.23) the result is

$$p = 1 - F_1(0) = 1 - 0{,}5 = 0{,}5. \tag{21.28}$$

If hypothesis H is invalid, then, in accordance with (21.24), the only known relative to magnitude p is

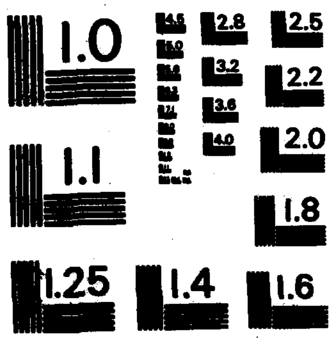
$$p \neq 0{,}5. \tag{21.29}$$

AD-A120 899 THEORY OF OPTIMUM RADIO RECEPTION METHODS IN RANDOM NOISE(U) FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OH L S GUTKIN 24 SEP 82 FTD-ID(RS)T-0784-82

7/7

UNCLASSIFIED F/G 9/4 NL

END



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Consequently, if hypothesis H is valid, then the magnitude z distribution turns out to be precisely known, in spite of the fact that distribution $F_1(y_i)$ of input data y_i is unknown. This stipulates the advisibility of preliminary transformation of input data $(y_1, \ldots, y_n)$ into magnitude z of the (21.26) form.

In order to substantiate the algorithm for further data z processing, we will examine a case when $n \gg 1$ (evidently, if there is no success in obtaining a satisfactory decision given a large sample size n, then it is all the more impossible to do so when n is small). Where $n \gg 1$, it is possible to assume that magnitude z has normal distribution, i. e.,

$$W(z)=\frac{1}{\sqrt{2\pi}\,\sigma_z}\exp\left\{-\frac{(z-\bar{z})^2}{2\sigma_z^2}\right\}. \tag{21.30}$$

It is not difficult to determine parameters $\bar{z}$ and σ_z^2 of this distribution.

Actually, random magnitude u may have only values 1 and 0 with probabilities p and $(1 - p)$, respectively. Therefore

$$\bar{u}=1p+0(1-p)=p. \tag{21.31}$$

But

$$\sigma_u^2=\overline{u^2}-(\bar{u})^2, \tag{21.32}$$

where $\overline{u^2}=1^2p+0^2(1-p)=p$.

Consequently,

$$\sigma_u^2=p-p^2=p(1-p). \tag{21.33}$$

But, it follows from (21.26) [considering independence of terms $u(y_i)$] that

$$\bar{z}=n\bar{u} \quad \text{and} \quad \sigma_z^2=n\sigma_u^2. \tag{21.34}$$

Therefore, the result is

$$\bar{z}=np \quad \text{and} \quad \sigma_z^2=np(1-p). \tag{21.35}$$

So, the decision to accept (or not accept) hypothesis H must be taken based on magnitude z, which has normal distribution (21.30) with parameters determined from relationships (21.35). If hypothesis H is valid, i. e., distribution $F(y_i)$ is even, then, in accordance with (21.28) and (21.35), the result is

$$\bar{z}=\frac{n}{2}, \quad \sigma_z=\frac{1}{2}\sqrt{n}. \tag{21.36}$$

It turns out that $\frac{\sigma_z}{\bar{z}}\to 0$ when $n \longrightarrow \infty$ and, consequently, it is possible /401 in the first approximation to assume that $z \approx n/2$. If hypothesis H is invalid, then, in accordance with (21.29), either $p < 1/2$ or $p > 1/2$ and, $z\approx\bar{z}>n/2$ or $z\approx\bar{z}<n/2$, respectively.

Therefore, it is natural to select that algorithm for testing the correctness of hypothesis H:

hypothesis H is valid if magnitude z is close enough to $n/2$, namely if

$$\frac{n}{2}-\xi\frac{\sqrt{n}}{2}<z\leqslant\frac{n}{2}+\xi\frac{\sqrt{n}}{2}. \tag{21.37}$$

Conversely, when magnitude z exceeds the bounds set by inequalities (21.37), hypothesis H is invalid and must be rejected. Here, ξ -- some fixed magnitude (for example, $\xi = 4$), selection of which must be made from permissible error decision probabilities.

Two types of errors are possible in this case: rejection of hypothesis H when it is valid and acceptance of hypothesis H when it is invalid. Let the probability of the first type error be α while that of the second type* be β. It follows from (21.37) that α is the probability that magnitude z will exceed the bounds (threshold) set by inequality (21.37), i. e., will deviate from

*In mathematical statistics, α is called criterion level of significance, while β is called criterion cardinality.

its expected value (mean value) by a magnitude exceeding $\xi\sigma_z$, where $\sigma_z = \sqrt{n}/2$ — magnitude z standard deviation. Since magnitude z distribution is normal, then it is evident that α will be very slight already when $\xi > 4$. Magnitude α decreases with a rise in ξ, but then magnitude β will rise, i. e., the probability that condition (21.37) is satisfied when hypothesis H is invalid. Therefore, selected magnitude ξ should not be too high, but such that probability α will equal the maximum permissible magnitude α_0.

From relationships (21.30), (21.35), (21.36), and (21.37), it is not difficult to obtain the following error probability formulas:

$$\alpha = 1 - 2\frac{1}{\sqrt{2\pi}}\int_0^{\xi} e^{-x^2/2}\,dx; \tag{21.38}$$

$$\beta = \frac{1}{\sqrt{2\pi}}\int_0^{\eta_1} e^{-x^2/2}\,dx - \frac{1}{\sqrt{2\pi}}\int_0^{\eta_2} e^{-x^2/2}\,dx, \tag{21.39}$$

where

$$\eta_1 = \frac{n(1/2-p)-\xi\sqrt{n}/2}{\sqrt{np(1-p)}}, \quad \eta_2 = \frac{n(1/2-p)+\xi\sqrt{n}/2}{\sqrt{np(1-p)}}.$$

It follows from formula (21.38) that, for given magnitude of probability $\alpha = \alpha_0$, the requisite value of threshold factor ξ may be determined precisely from probability integral tables. Where $\alpha \leqslant 10^{-4}$, it is possible to assume /402

$$\xi \approx \sqrt{2\ln\frac{1}{\alpha}}. \tag{21.40}$$

It follows from (21.39) that computation of probability β (acceptance of hypothesis H when it is invalid) requires know[illegible]ng (along with already-selected magnitude ξ) parameter p of a Poisson dis[illegible]ion. But, in cases when hypothesis H is invalid, i. e., unknown distribution F([illegible]) [illegible] not have the property of parity, parameter p is unknown; the only known is that p differs from 0.5 [and this difference is all the greater, the greater the distribution $F(y_i)$ asymmetry]. Therefore, it is impossible precisely to determine the probability β magnitude. However, it follows from this formula that, when $p \neq 1/2$ and $n \to \infty$, probability

β will strive towards zero. Therefore, one may assert that a slight probability β value will be insured, given a large-enough sample size n. (Evidently, the more p differs from 1/2, the smaller the required sample size n, for a given β magnitude).

Thus, a decision rule based on relationships (21.25), (21.26), and (21.37) insures (where $n \gg 1$) receipt of a slight probability error of the first kind and an error probability of the second kind asymptotically (where $n \longrightarrow \infty$) striving towards zero. Therefore, such a decision rule may be considered acceptable under the aforementioned conditions where a priori information is lacking almost completely.

We will examine the following problem as an example of the use of the given decision rule for receiver system design.

Consider binary detection of a signal on a noise background based on analysis of input realization y(t). The only a priori data on the signal and noise is the following.

1. It is known that, if samples $(y_1, \ldots, y_i, \ldots, y_n)$ with interval $\Delta t \geqslant \Delta t_{\text{мин}} \ll T$ (where T — time allocated for detection) are taken from realization y(t), then sample values will be statistically independent and have identical distribution $F(y_i)$.

2. Nothing is known about distribution $F(y_i)$ except that, when there is no signal, it may be considered even, while parity is disrupted when there is a signal.*

Since

$$n = \frac{T}{\Delta t_{\text{мин}}} \gg 1, \qquad (21.41)$$

then the aforementioned decision rule may be used. To do so, input data y(t) are quantized with respect to time with interval $\Delta t = \Delta t_{\text{мин}}$ [for formation of sample

*In other words, when there is no signal, positive and negative sample values y_i are equally probable, while signal appearance renders them non-equally possible.

$(y_1, \ldots, y_n)$] and with respect to magnitude in the zero threshold, /403 resulting in formation of quantized sample values $u(y_i)$. Then, these data are added [in accordance with law (21.26)] and the resultant sum is compared with thresholds

$$z_{\text{мин}} = \frac{n}{2} - \xi \frac{\sqrt{n}}{2} \quad \text{and} \quad z_{\text{макс}} = \frac{n}{2} + \xi \frac{\sqrt{n}}{2}.$$

If magnitude z does not exceed these bounds, then a decision is taken concerning parity of the distribution giving rise to realization y(t), i. e., concerning signal absence. Conversely, i. e., when any threshold is exceeded, the decision taken is that a signal is present. Here, α is the false-alarm probability, while β is miss probability, i. e.,

$$\alpha = P_{\text{лт}}, \quad \beta = P_{\text{пр}}. \tag{21.42}$$

Therefore, being given the permissible probability $P_{\text{лт}}$ value, it is possible to determine requisite threshold factor ξ value from formula (21.38) or (where $P_{\text{лт}} \leqslant 10^{-4}$) from formula (21.40). The decision rule (detector operating algorithm) selected guarantees the given slight probability $P_{\text{лт}}$ magnitude and miss probability $P_{\text{пр}}$ asymptotically (where $n \rightarrow \infty$) striving towards zero.

We now will examine another nonparametric method of statistics based on use of the congruence criterion. The quantitative measure characterizing the divergence ("disagreement") between sample value probabilities F_1 ( ) and F_2 ( ) is called the congruence criterion $\Delta(F_1, F_2)$. Depending on the nature of the problem posed, both distributions may be empirical, or one of them may be empirical and the other hypothetical (approximate).

We will examine the problem of testing hypothesis H concerning the affiliation of two samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$ with independent elements with the same probability distribution against alternative hypothesis K that these samples are affiliated with different distributions.

Empirical distributions $F_x(z)$ and $F_y(z)$ built from samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$, respectively, play the role of distributions F_1 and F_2 in

this case. By definition, the empirical function of a distribution corresponding to sample $(x_1, \ldots, x_n)$ equals

$$F(x)=\frac{\nu_1(x)}{n}, \tag{21.43}$$

where $\nu_1(x)$ -- number of sample values not exceeding some threshold x. For computation of $\nu_1(x)$, sample values $(x_1, \ldots, x_n)$ are presorted into ascending magnitudes, i. e., in the form of a so-called variational series.

The empirical funtion of distribution F(y) corresponding to sample $(y_1, \ldots, y_n)$ is determined in an analogous manner:

$$F(y)=\frac{\nu_2(y)}{m}. \tag{21.44}$$

Then /404

$$\left.\begin{aligned}F_x(z)&=F(x)_{x=z},\\ F_y(z)&=F(y)_{y=z}.\end{aligned}\right\} \tag{21.45}$$

So, to test hypothesis H that distributions $F_x(z)$ and $F_y(z)$ coincide against alternative K that they do not coincide requires preliminary selection of the congruence criterion type, i. e., the type of function (or functional) $\Delta(F_x, F_y)$ characterizing the degree of divergence between F_x and F_y. Different congruence criteria, each having inherent advantages and disadvantages, correspond to the different types of this function (functional). We will examine as our example two congruence criterion types:

The first congruence criterion has the form

$$\Delta=\max_z|F_x(z)-F_y(z)|. \tag{21.46}$$

As N. V. Smirnov demonstrated [131], when $n\to\infty$

$$P\left\{\left(\frac{1}{n}+\frac{1}{m}\right)^{-1/2}\Delta>z\right\}=2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2z^2},\quad z>0, \tag{21.47}$$

where P { } -- probability of the event enclosed in brackets. The decision rule may be selected in the following form: hypothesis H is true [samples $x_1, \ldots, x_n$) and ($y_1, \ldots, y_n$) will be affiliated with the same distribution] if

$$\Delta \leqslant \Delta_\alpha. \tag{21.48}$$

If inequality (21.48) is not satisfied, then hypothesis H is untrue, i. e., alternative hypothesis K is valid. Here, Δ_α -- some threshold selected from the permissible value of probability α of rejecting hypothesis H when it is true.

It is possible on the basis of relationships (21.47) and (21.48) to find the link between probability α and threshold Δ_α, making it possible to select magnitude Δ_α in accordance with the given (permissible) probability α value. Actually, it follows from determination of probability α that

$$\alpha = P\{\Delta > \Delta_\alpha\}, \tag{21.49}$$

given that hypothesis H is true. But, if hypothesis H is true, then the distribution of magnitude Δ is subordinate to law (21.47). Therefore, formula (21.47) may be used for determination of probability $\alpha = P\{\Delta > \Delta_\alpha\}$. Evidently, it is possible to write this formula in the following way as well:

$$P\left\{\Delta > \left(\frac{1}{n} + \frac{1}{m}\right)^{1/2} z\right\} = 2\sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 z^2}.$$

Substituting this expression into (21.49), we obtain

$$\alpha = 2\sum_{k=1}^{\infty} (-1)^{k-1} \exp\left[-2k^2\left(\frac{1}{n} + \frac{1}{m}\right)^{-1} \Delta_\alpha^2\right].$$

Given slight values of α, it is possible to restrict ourselves to the first term of the sum, i. e., to assume /405

$$\alpha \approx 2\exp\left[-\left(\frac{1}{n} + \frac{1}{m}\right)^{-1} \Delta_\alpha^2\right].$$

hence, the result is

$$\Delta_\alpha \approx \sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{1}{2}\ln\frac{2}{\alpha}}.$$

Another congruence criterion (named the Wilcoxon criterion for its author) is based on a count of the number of inversions. In the problem formulated above, when this criterion is used, both samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$ are in the form of one (common) variational series, for example:

$$x_2, x_5, y_4, y_1, x_3, y_3, x_1, \ldots$$

If, in this series, k elements of sample $(y_1, \ldots, y_m)$ precede given x_i, then k inversions occur. The total number of inversions Δ equals the sum of the inversions formed by all elements of the first sample with the elements of the second.

Where $m + n \geqslant 20$ and $m > 3$, it is possible approximately to consider that magnitude Δ has normal distribution with parameters

$$\overline{\Delta}=\frac{mn}{2}, \quad \sigma_\Delta^2=\frac{mn}{12}(m+n+1). \tag{21.50}$$

The decision rule has the following form

if $$\Delta < \alpha, \tag{21.51}$$

then hypothesis H is accepted, while, conversely, hypothesis K is accepted.

It is possible to use relationships (21.50) and (21.51) to find the link between probability α of hypothesis H rejection when it is true and threshold Δ_α:

$$\Delta_\alpha=\frac{mn}{2}+x_\alpha\sqrt{\frac{mn}{12}(m+n+1)}. \tag{21.52}$$

where x_α is determined from relationship

$$\frac{1}{\sqrt{2\pi}}\int_{x_\alpha}^{\infty} e^{-z^2/2}\,dz=\alpha. \tag{21.53}$$

When $\alpha \leqslant 10^{-4}$, it is possible to assume that

$$x_\alpha \approx \sqrt{2 \ln \frac{1}{\alpha}}. \quad (21.54)$$

As is evident from the examples presented, congruence criteria make it possible (given large enough samples) precisely to determine probability α of an error of the first type (rejection of hypothesis H when it is true), but do not /406 make it possible to determine probability β of an error of the second type (acceptance of hypothesis H when it is untrue).

It is possible to solve, for example, the following problem of detecting a signal on a noise background using the aforementioned methods based on the congruence criterion.

The requirement is to establish whether or not a signal is present in realization $y^{(1)}(t)$ given the presence of additional (pattern) realization $y^{(2)}(t)$. It is known that there cannot be a signal in realization $y^{(2)}(t)$ and the statistical characteristics of the noise in realizations $y^{(1)}(t)$ and $y^{(2)}(t)$ are identical. The duration of the first realization is T_1, while the duration of the signal appearing in it also is T_1. The duration of the second realization is T_2. Durations T_1 and T_2 are so long that it is possible to obtain from realizations $y^{(1)}(t)$ and $y^{(2)}(t)$ n and m independent sample values, respectively, where

$$n \gg 1 \text{ and } m \gg 1. \quad (21.55)$$

The fact that the laws of distribution of the sample values within each realization are identical is assumed to be known. Here, uniform (integral) distribution $F_1(z)$ completely characterizes the law of distribution of the first sample, while uniform distribution $F_2(z)$ completely characterizes the second sample. If there is no signal in the first sample, then distribution $F_1(z)$ coincides with distribution $F_2(z)$. Appearance of a signal (in the first sample) causes $F_1(z)$ to deviate from $F_2(z)$.

Consequently, testing the hypothesis H concerning the coincidence of distributions F_1 and F_2 is fully equivalent to testing the hypothesis concerning signal absence. Hypothesis K concerning signal presence is the alternative hypothesis. Hence,

it follows that, under the given conditions, any of the congruence criteria described above may be used. Here, samples obtained from realizations $y^{(1)}(t)$ and $y^{(2)}(t)$ play the role of samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$, respectively, and α and β are false-alarm and signal miss probabilities, respectively.

These examples clearly illustrate the following special features of nonparametric methods of statistics. These methods make it possible to choose a receiver operating algorithm insuring precise equality of false-alarm probability $P_{\pi\tau}$ of a given (permissible) magnitude, even for very-slight a priori information concerning the laws of signal-plus-noise distribution. However, the paucity of this information leads to the fact that there is no success in determining the precise miss probability $P_{\pi\pi}$ value; in the best case, it is possible to obtain only a qualitative guarantee of the smallness of this probability when $n \longrightarrow \infty$. In addition, the aforementioned nonparametric methods have the following serious deficiencies:

1. There is a requirement for a large number of statistically-independent sample values, which leads to the necessity to select a large enough time interval between adjacent realization sample values and, correspondingly, to lose some of the usable information contained in these realizations.

2. Each of the described methods is valid only for a relatively-narrow class of problems. For example, there is the requirement that the law of distribution /407 be even when there is no signal or that an additional ("pattern") sample exist, which does not a fortiori contain a signal. In addition, there is a requirement that the law of distribution of all sample values (within a given sample) be identical.

3. In the general case, there is success only in finding acceptable algorithms $\Gamma(y)$, rather than the best-possible ones (within the framework of the given quality criterion).

However, the theory of nonparametric methods of statistics is in a state of rapid development and it may turn out to be useful for a number of problems with very slight a priori information.

21.4 Use of the Theory of Stochastic Approximation

Robbins and Monro [192] for the first time proposed a method of stochastic approximation to find the estimate of the root of a regression equation.

Regression is one of the characteristics of the statistical link between two random magnitudes x and y and, by definition, equals the expected value of one of these magnitudes, given a fixed value of the other, i. e., $M_x(y)$ -- regression of y to x; $M_y(x)$ -- regression of x to y.

Dependence $M_x(y) = m(x)$ (or dependence of $M_y(x)$ on y) is called a regression function (curve), while the equation

$$m(x) = 0 \tag{21.56}$$

is a regression equation. Function m(x) has a very-valuable property: if one accepts as the estimate y^* of dependence y on x functional dependence of the type

$$y^* = m(x) \tag{21.57}$$

then the standard deviation of y^* from y will be minimal.

Robbins and Monro proposed an iterative procedure of so-called stochastic approximation for finding estimate x_o^* of root x_o of regression equation (21.56) given sample $(y_1, \ldots, y_k, \ldots, y_n)$ with elements having some conditional distribution $P_x(y_k)$ [in the supposition that equation (21.56) has a total of one root]. Here, functions $P_x(y_k)$ and m(x) are considered unknown so that estimate x_o^* of the magnitude of the equation (21.56) root must be given only based on sample $(y_1, \ldots, y_k, \ldots, y_n)$ for each element y_k of which

$$M_{x_k^*}(y_k) = m(x_k^*).$$

Estimate x_o^* of the needed root will be found by means of the n-th use of recurrent relationship

$$x_{k+1}^* = x_k^* + a_k y_k, \tag{21.58}$$

where k = 1, . . ., n. Here, a random constant number may be selected as $x_1^* = x_1$.

If coefficients a_k satisfy the conditions+ /408

$$\sum_{k=1}^{\infty} a_k = \infty, \quad \sum_{k=1}^{\infty} a_k^2 < \infty, \tag{21.59}$$

then, given an unrestricted increase in sample size ($n \longrightarrow \infty$), estimate x_{n+1}^* will diverge from the probability of unity to needed root x_0, i. e.,

$$P\left\{\lim_{n\to\infty} x_{n+1}^* = x_0\right\} = 1. \tag{21.60}$$

Conditions (21.59) are satisfied in particular if

$$a_n = 1/n. \tag{21.61}$$

Later, Kiefer and Wolfowitz [175] proposed a corresponding procedure of stochastic approximation for finding the extremum of regression function m(x) (in the supposition that there is only one extremum).

Ya. Z. Tsypkin in his works [132, 172, 198] demonstrated that stochastic approximation algorithms successfully may be used in design of various information retrieval systems and control systems. We will shift to a brief exposition of the main content of Tsypkin's work relative to radio reception system design problems.

Let the following population be subjected to system design (optimum selection)

$$\vec{C} = \{c_1, c_2, \ldots, c_N\} \tag{21.62}$$

of receiver parameters c_1, c_2, . . ., c_N, while this receiver's quality index may be represented in the form

$$\mathcal{J} = \mathcal{J}(\vec{C}) = \int_{A_y} Q(y, \vec{C})\, P(y)\, dy = M\{Q(y, \vec{C})\}, \tag{21.63}$$

+In addition to restrictions (21.59) on coefficients a_k, some (usually insignificant) restrictions also are placed on the values of conditional (for a given x) expected value and variance of random variable y (see [195], for example).

where function $Q(y, \vec{C})$ considers the dependence of the quality index both on vector $\vec{C}$ and on input realization $y = y(t)$. [In the case of discrete sample $y = (y_1, \ldots, y_n)$].

Optimum vector $\vec{C}$ value $\vec{C}_{\text{кр}}$ is considered that value to which the magnitude $\mathcal{J}$ extremum (maximum or minimum, depending on the type of problem being solved) corresponds. For simplicity, in future (but without disrupting the generality of the results) it is possible to assume that the minimum is this extremum.

There is success in reducing the majority of quality indices used at the present time to the (21.63) form, particularly average risk R. If function /409 $\mathcal{J}(\vec{C})$ is differentiated for $\vec{C}$, then the requisite condition of the magnitude $\mathcal{J}$ extremum has the following form:

$$\nabla \mathcal{J}(\vec{C}) = 0, \tag{21.64}$$

where

$$\nabla \mathcal{J}(\vec{C}) = \left\{\frac{\partial \mathcal{J}}{\partial c_1}, \ldots, \frac{\partial \mathcal{J}}{\partial c_N}\right\} \tag{21.65}$$

-- gradient of vector $\vec{C}$. It follows from (21.63) and (21.64) that the extremum condition also may be written:

$$M\left\{\nabla_c Q(y, \vec{C})\right\} = 0, \tag{21.66}$$

where

$$\nabla_c Q(y, \vec{C}) = \left\{\frac{\partial Q}{\partial c_1}, \ldots, \frac{\partial Q}{\partial c_N}\right\} \tag{21.67}$$

-- gradient of vector $Q(y, \vec{C})$ with respect to $\vec{C}$.

Vector $\vec{C}$ optimum value $\vec{C}_{\text{кр}}$, i. e., the value satisfying equation (21.66), may be found from the following iterative algorithm, which is one of the variations of the stochastic approximation algorithm:

$$\vec{C}[n] = \vec{C}[n-1] - A[n] \nabla_c Q(y[n], \vec{C}[n-1]), \tag{21.68}$$

where $\vec{C}[n]$ and $\vec{C}[n-1]$ -- needed vector values in the n-th and (n — 1)-th approximation step. Initial vector $\vec{C}$ value $\vec{C}[1]$ may be random; A[n] -- diagonal N x N matrix of the type

$$A[n]=\begin{Vmatrix} a_1[n], 0, \ldots, 0 \\ 0, a_2[n], \ldots, 0 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ 0, 0, \ldots, 0, a_N[n] \end{Vmatrix}. \tag{21.69}$$

Simpler is the particular case when, instead of (21.69), the following is selected

$$A[n]=Ia[n], \tag{21.70}$$

where I — identity matrix.

Evidently, case (21.69) differs from (21.70) in that the values of coefficients $a_1[n], \ldots, a_N[n]$ selected for different vector $\vec{C}$ components $c_1, \ldots, c_N$ are not identical.

Algorithm (21.68) differs from the conventional algorithm of an extremum gradient search in that gradient $\nabla_c Q$ will depend on random variable y and, consequently, is random. Therefore, special substantiation is required for the convergence (when $n \longrightarrow \infty$) of magnitude $\vec{C}[n]$ to the equation (21.66) root, i. e., to extreme value $\vec{C}_{\text{кр}}$.

Analogous to the case where coefficient a_n selection in accordance with /410 condition (21.59) is required to insure convergence of algorithm (21.58) in a Robbins and Monro problem, corresponding selection of coefficients $a_i[n]$ of matrix A[n] is required for algorithm (21.68) convergence.* But, when solving practical problems, there is a requirement that not only the algorithm will converge when $n \longrightarrow \infty$, but that it will do so rapidly enough. Therefore, substantiation of the best selection of the type of matrix A[n] or [in the case of (21.70)] dependence

*In addition, several additional conditions analogous to those indicated in the page 584 note must be satisfied.

a[n] is the central task in the theory of stochastic approximation. This task still is far from satisfactory solution, but some useful results already have been obtained. For example, sufficient conditions of convergence with a probability of unity for algorithm (21.68) [for the (21.70) case] is presented in [132, page 72]. They include inequalities (21.59 and several additional conditions.

So, if coefficient A[n] in algorithm (21.68) satisfies the convergence conditions, then this algorithm makes it possible, when $n \longrightarrow \infty$, to obtain optimum value $\vec{C}_{op}$ of the system parameter vector. In future, we will assume for simplicity that A[n] is selected in accordance with (21.70), i. e.,

$$\vec{C}[n]=\vec{C}[n-1]-Ia[n]\nabla_c Q(y[n], \vec{C}[n-1]), \qquad (21.71)$$

where $\vec{C}[1]$ -- random vector $\vec{C}$ initial value.

In a number of problems, there is no success in computing gradient $\nabla_c Q(y, \vec{C})$ [because of the nondifferentiability of function $Q(y, \vec{C})$ with respect to $\vec{C}$, for example]. In such cases, gradient $\nabla_c Q$ may be found (measured) approximately, for example, using the following retrieval method.

We assume that

$$\nabla_c Q(y, \vec{C}) \approx \tilde{\nabla}_c Q(y, \vec{C}, b), \qquad (21.72)$$

where

$$\tilde{\nabla}_c Q(y, \vec{C}, b)=\frac{Q_+(y, \vec{C}, b)-Q_-(y, \vec{C}, b)}{2b}, \qquad (21.73)$$

/411

$$\left.\begin{aligned} Q_+(y, \vec{C}, b) &= \{Q(y, \vec{C}+be_1), \ldots, Q(y, \vec{C}+be_N)\}\cdot \\ Q_-(y, \vec{C}, b) &= \{Q(y, \vec{C}-be_1), \ldots, Q(y, \vec{C}-be_N)\}\cdot \end{aligned}\right\} \qquad (21.74)$$

Here, b -- scalar, e_i(i = 1, . . ., N) -- basis vectors. In the simplest case

$$e_1=(1, 0, \ldots, 0); \quad e_2=(0, 1, 0, \ldots, 0), \ldots, e_N=(0, 0, \ldots, 0, 1). \qquad (21.75)$$

Magnitude $\tilde{\nabla}_c Q(y, \vec{C}, b)$ may be determined, for example, with the aid of a synchronous detector (Figure 21.3) in which reference voltage has the form of

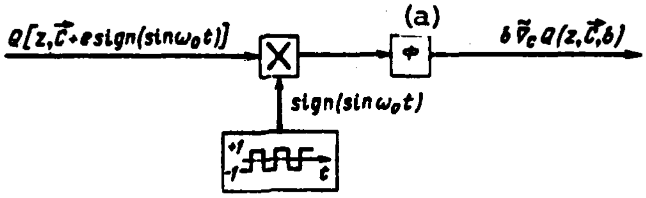


Figure 21.3. (a) -- F [filter].

a rectangular identity function sign $(\sin \omega_0 t)$. Filter Φ at synchronous detector output suppresses rf components. Here, the structural schematic of a discrete servomechanism realizing algorithm (21.71) [in which gradient $\nabla_c Q(y, \vec{C})$ is computed

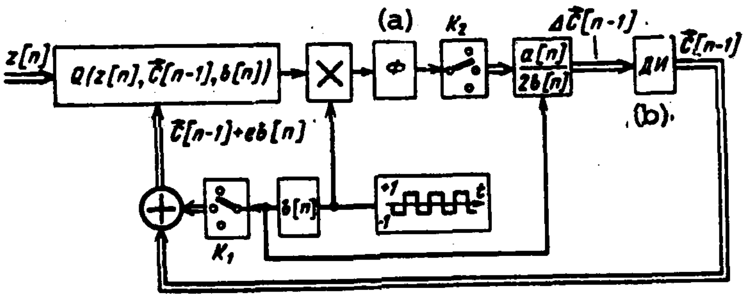


Figure 21.4. (a) -- F [filter]; (b) -- DI [discrete integrator].

from formula (21.72)] has the form depicted in Figure 21.4. Vector links are designated by double lines; K_1 and K_2 -- commutators; Φ -- filter; *ДИ* -- discrete integrator (digrator), which converts difference $\vec{\Delta C}[n-1] = \vec{C}[n] - \vec{C}[n-1]$ into $\vec{C}[n-1]$. It will comprise delay element (by one step) *ЭЗ* (Figure 21.5) enclosed by unitary positive feedback.

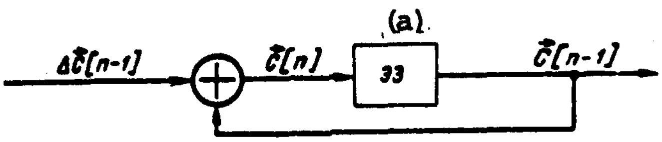


Figure 21.5. (a) -- EZ [delay element].

When the retrieval method is used to determine the gradient, i. e., in accordance with expressions (21.72)—(21.74), optimum vector $\vec{C}$ value computational algorithm (21.71) takes the following form:

$$\vec{C}[n]=\vec{C}[n-1]-Ia[n]\tilde{\nabla}_{c}Q(y[n],\vec{C}[n-1]). \quad (21.76)$$

In a number of cases, instead of discrete algorithm (21.71), it is more /412 convenient to use an analog algorithm of the type

$$\frac{d\vec{C}(t)}{dt}=-a(t)\nabla_{c}Q(y(t),\vec{C}(t)). \quad (21.77)$$

When $n \longrightarrow \infty$, solution of this equation will strive towards optimum value $\vec{C}_{\text{мр}}$ of the system parameter vector, given any initial value of vector $\vec{C}=\vec{C}(0)$. Coefficient a(t) is a continuous analog of discrete function a[n] and must be selected from the condition of the fastest vector $\vec{C}(t)$ convergence to $\vec{C}_{\text{мр}}$. The discrete algorithm is convenient for realization in digital computers, while the analog algorithm is convenient for analog computers.

We will examine the following two examples of the multitude presented during explanation of the methodology from [132, 198].

Example 1. Use in design of a Wiener optimum linear filter system.

The problem of linear filtration in the Wiener formulation was examined in § 2.2. It will consist of the determination of impulse response $\eta(t)$ of a linear stationary filter providing the minimum message reproduction mean-square error, i. e., the minimum of magnitude

$$\mathcal{F}=M[h(t)-\gamma(t)]^{2}, \quad (21.78)$$

where $h(t)=\mathcal{D}^{*}(p)u_{c}(t)$ -- desired effect at filter output, while

$$\gamma(t)=\int_{0}^{\infty}y(t-\tau)\eta(\tau)d\tau \quad (21.79)$$

-- actual oscillation at filter output;

$$y(t)=u_c(t)+u_{ш}(t)$$

-- sum of random signal and noise at filter input.

Formula (21.79) applies to a case when signal-plus-noise y(t) reaches filter input at moment $t_0 = -\infty$. If $t_0 = 0$, then, instead of (21.79), one should assume that

$$\gamma(t)=\int_0^t y(t-\tau)\,\eta(\tau)\,d\tau. \tag{21.79a}$$

Here, formula (21.78) takes the following form:

$$\mathcal{F}=M\,[h(t)-\int_0^t \eta(\tau)\,y(t-\tau)d\tau]^2. \tag{21.80}$$

It was noted in § 2.2 that finding the optimum impulse response $\eta_{опт}(t)$ providing the minimum of functional $\mathcal{F}$ requires knowledge of signal and noise correlation functions. If these functions are unknown, then it is possible to use the stochastic approximation methodology presented above. To do so, the needed impulse reaction is represented in the form

$$\eta(t)=c_1\varphi_1(t)+c_2\varphi_2(t)+\ldots+c_N\varphi_N(t), \tag{21.81}$$

where $\varphi_1(t), \ldots, \varphi_N(t)$ -- known (preselected) linearly-independent time functions. Here, the problem of finding optimum impulse response $\eta(t)$ boils down to finding optimum parameter vector $\vec{C}$.

Expression (21.81) may be written in abbreviated form: /413

$$\eta(t)=\vec{C}^T\,\vec{\varphi}(t), \tag{21.82}$$

$\vec{C}^T$ -- transposed vector $\vec{C}$ (vector-line), while $\vec{\varphi}(t)$ -- N-dimensional vector, whose components are function $\varphi_1(t), \ldots, \varphi_N(t)$. Substituting (21.82) into (21.80), we obtain

$$\mathcal{F}(\vec{C})=M\{[h(t)-\vec{C}^T\vec{\varphi}(t)]^2\}, \tag{21.83}$$

where

$$\vec{\psi}(t)=\int_0^t \vec{\varphi}(\tau)\, y(t-\tau)\, d\tau. \tag{21.84}$$

Comparing expressions (21.83) and (21.84) with (21.63), it is easy to become convinced that, in this case, the role of function $Q(y, \vec{C})$ is played by function

$$Q(y,\vec{C})=\left[h(t)-\vec{C}^T\vec{\psi}(t)\right]^2=\left[h(t)-\vec{C}^T\int_0^t \varphi(\tau)\, y(t-\tau)\, d\tau\right]^2. \tag{21.85}$$

Therefore, it is possible to use algorithm (21.77) to find optimum vector $\vec{C}$ value $\vec{C}_{\text{нр}}$, i. e., to find $\vec{C}_{\text{нр}}$ as a steady-state (where $n \longrightarrow \infty$) value of solution $\vec{C}(t)$ of the differential equation

$$\frac{d\vec{C}(t)}{dt}=-a(t)\,\nabla_c Q\left(y(t),\vec{C}(t)\right). \tag{21.86}$$

But, it follows from (21.67), (21.84), and (21.85) that

$$\nabla_c Q\left(y(t),\vec{C}(t)\right)=-2\left[h(t)-\vec{C}^T\vec{\psi}(t)\right]\vec{\psi}(t).$$

Therefore, algorithm (21.86) has the following form:

$$\frac{d\vec{C}(t)}{dt}=2a(t)\left[h(t)-\vec{C}^T\vec{\psi}(t)\right]\vec{\psi}(t). \tag{21.87}$$

As noted above, the solution of this equation will converge (when $t \longrightarrow \infty$) towards needed optimum value $\vec{C}_{\text{нр}}$ for any value of initial condition $\vec{C}(0) = \vec{C}_0$.

It follows from relationship (21.87) that it is necessary to assume desired filter output reaction $h(t) = \mathscr{D}(p)\, u_c(t)$ to determine optimum value $\vec{C}_{\text{нр}}$ of parameter vector $\vec{C}$. But, to do so, one must have a realization of signal $u_c(t)$ forming part of signal-plus-noise $y(t)$. In the filter operating mode, it is impossible in principle to provide this. Therefore, when algorithm (21.87) is used, it is necessary to supply to filter input special pattern realizations

$$y_0(t)=u_{c\,0}(t)+u_{ш\,0}(t),$$

which allow shaping of desired output realizations

$$h(t)=\mathscr{D}(p)\,u_{c\,0}(t), \tag{21.88}$$

where $\mathscr{D}(p)$ -- known (given) linear operator.

After the pattern process is completed, i. e., the estimate of the optimum parameter vector $\vec{C}_{HP}$ value is found, the designed filter may be switched to the operating mode, in which operating realizations y(t), which contain unknown /414 signal $u_c(t)$, reach its input. Strictly speaking, $\vec{C}$ will converge to $\vec{C}_{HP}$ only when $t \longrightarrow \infty$, i. e., the pattern process must last an infinitely-long time. However, if the vector $\vec{C}(t)$ convergence to $\vec{C}_{HP}$ is good enough (fast enough), then it is possible to assume, with accuracy satisfactory for practical purposes, that pattern time is finite.

To illustrate these suppositions, we will examine the simplest case when a filter impulse response is sought in the form

$$\eta(t)=c_1\,\varphi_1(t). \tag{21.89}$$

This denotes that the impulse response shape is considered given and unknown, i. e., only scale factor c_1 is subject to system design.

Here, vector differential equation (21.87) is converted into a conventional (scalar) equation

$$\left.\begin{aligned}
&\frac{dc}{dt}=2a(t)\,[h(t)-c_1\,\psi_1(t)]\,\psi_1(t),\\
&\text{where}\\
&h(t)=\mathscr{D}(p)\,u_{c\,0}(t);\\
&\psi_1(t)=\int_0^t \varphi_1(\tau)\,y_0(t-\tau)\,d\tau;\\
&y(t)=u_{c\,0}(t)+u_{ш\,0}(t).
\end{aligned}\right\} \tag{21.90}$$

The Figure 21.6 structural schematic corresponds to these expressions. In this figure, Φ_1 -- linear filter with impulse response $\varphi_1(t)$; Φ_2 -- linear

filter with given operator transfer constant $\Phi(p)$; $И$ -- integrator; $П_1, \ldots, П_5$ -- switches from the pattern mode (position 1) to the operating mode (position 2).

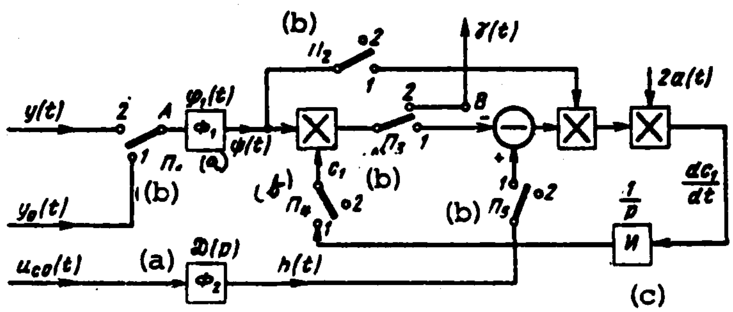


Figure 21.6. (a) -- F [filter]; (b) -- P [switch]; (c) -- I [integrator].

Determinate functions $\varphi_1(t)$ and $a(t)$ are assumed to be selected prior to beginning the pattern; u_{co}, $u_c(t)$, $u_{ш0}(t)$, and $u_ш(t)$ -- stationary random processes (with zero mean values). The statistical characteristics (correlation functions) of these processes may be unknown; it is important only that these characteristics do not change during the transition from $u_{ш0}(t)$ and $u_{co}(t)$ to $u_ш(t)$ and $u_c(t)$.

During the pattern process, realizations $y_0(t)$ of the pattern random process reach filter input. Usable signal realizations $u_{co}(t)$ included in this process are known here and are used simultaneously to form the desired effect $h(t)$ at filter output. Evidently, function $h(t)$ plays the role of "teacher:" it "indicates" to the filter which realization $\gamma(t)$ must be at output if there is no noise. During the pattern process, needed parameter c_1 is a time function and, in /415 the steady-state mode (where $n \longrightarrow \infty$), it must equal constant magnitude $c_{нр}$. Therefore, the moment function $c_1(t)$ achieves an approximately stationary value (with error not exceeding 1%, for example) may serve as the criterion for practical cessation of the pattern process.

Upon cessation of the pattern process, the designed filter's impulse response must equal [in accordance with (21.89)]

$$\eta(t) = c_1 \varphi_1(t). \quad (21.91)$$

where $c_1 \approx c_{1\,нр}$.

Since filter Φ_1 (Figure 21.6) has impulse response $\dot{\varphi}_1(t)$, section AB of the Figure 21.6 circuit has (upon cessation of the pattern) a (21.91)-type impulse response. Consequently, in the operating mode, the circuit section comprising filter Φ_1 and the constant factor c_1 multiplier must play the role of needed optimum linear filter. Therefore, in the operating mode, realizations y(t) are supplied to filter Φ_1 input, while result $\gamma(t)$ of the reproduction of signal $u_c(t)$ is picked off the multiplier output at point B.

Example 2. Use in signal detector system design.

Supplied to detector input is noise $u_{\text{ш}}(t)$ or signal-plus-noise

$$y(t) = u_c(t; \alpha; \ldots; \alpha_m) + u_{\text{ш}}(t). \tag{21.92}$$

Let the following composite error probability be the optimized quality index

$$P_{\text{ош}} = P(0) P_{\text{лт}} + P(C) P_{\text{пр}}, \tag{21.93}$$

i. e., a priori probabilities P(O) and P(C) of signal absence and presence are known. However, the statistical characteristics of the signal (when it is present) and the noise are unknown. Therefore, conditional distributions $P_{x_0}(y)$ and $P_{x_1}(y)$ also are unknown.

Under these conditions, given the presence of pattern realizations $y_0(t)$, optimum detector system design is possible based on pattern algorithm (21.68). It order for it to be used, quality index (21.93) must be reduced to the (21.63) form. This may be done by considering the following special binary detection features examined in § 14.3.

During binary detection, the decision rule may be written in the following form:

$$\gamma = \Gamma(y) = \begin{cases} 1, \text{ if } y \in A; \\ 0, \text{ if } y \in B, \end{cases} \tag{21.94}$$

where unity denotes a "yes" decision (signal) and zero a "no" decision (no signal);

y = y(t) -- input realization observed during time T. The requisite (ideal) decisions here are:

$$\gamma_0 = \begin{cases} 1, & \text{if } u_c(t) \neq 0, \\ 0, & \text{if } u_c(t) \equiv 0 \end{cases} \tag{21.95}$$

(here $u_c(t) \neq 0$ denotes signal presence, while $u_c(t) \equiv 0$ denotes signal absence).

As noted in § 14.2, optimum detector system design will boil down in essence to seeking the optimum boundary between osculating regions A and B of decisions "1" and "0," respectively. Algorithm (21.68) is applicable only in those cases when only population $\vec{C}$ of parameters of a system of known structure is subjected to system design. Therefore, the recommendation in [132] is to find the boundary between regions A and B in the form

$$f(y, \vec{C}) = c_1\varphi_1(y) + \ldots + c_N\varphi_N(y) = \vec{C}^T\vec{\varphi}(y) = 0. \tag{21.96}$$

where $\varphi_1(y), \ldots, \varphi_N(y)$ -- known (preselected) linearly-independent functions or, in the more general case, known operators (processing algorithms) of realizations y(t).

This denotes that the decision rule is sought in the form $\gamma = \Gamma(y, C)$, /416 i. e., has the following form:

if

$$\vec{C}^T\vec{\varphi}(y) > 0, \tag{21.97}$$

then realization y(t) belongs to region A and, consequently, the decision taken must be $\gamma = 1$;

if condition (21.97) is not satisfied, i. e.,

$$\vec{C}^T\vec{\varphi}(y) < 0,$$

then realization y(t) falls in region B and the decision taken is $\gamma = 0$.

In light of this, quality index (21.93) may be represented in the following form:

$$\mathcal{F}(\vec{C}) = P_{om}(\vec{C}) = P\{\gamma(y, \vec{C}) \dotplus \gamma_0 = 1\}, \tag{21.98}$$

where P { } denotes that, as usual, the probability of the occurence of the event indicated in the brackets.

Actually, it follows from (21.94) and (21.95) that both γ and γ_o may receive only values 1 and 0. Therefore, the sum ($\gamma + \gamma_0$) may equal unity only in one of the following incompatible cases:

a) either $\gamma = 1$, $\gamma_o = 0$; (21.99a)
b) or $\gamma = 0$, $\gamma_o = 1$ (21.99b)

But, it follows from (21.95) that $\gamma_o = 0$ when and only when $u_c(t) \equiv 0$, while $\gamma_o = 1$ when and only when $u_c(t) \doteq 0$. Therefore, conditions (21.99a) and (21.99b) also may be written in the following form:

a) either $\gamma = 1$, $u_c(t) \equiv 0$ (21.100a)
b) or $\gamma = 0$, $u_c(t) \neq 0$ (21.100b)

But, the probability of the first of these events equals $P(0)\,P_{\text{лт}}$, while the probability of the second event equals $P(C)\,P_{\text{пр}}$. Therefore, the probability of the occurence of even one of the two incompatible events equals

$$P(0)\,P_{\text{лт}} + P(C)\,P_{\text{пр}} = P_{om}.$$

Consequently, quality index (21.98) actually completely coincides with initial index (21.93).

We will designate

$$\xi(y, \vec{C}) = \gamma(y, \vec{C}) + \gamma_0. \tag{21.101}$$

Then, from (21.98), we have

$$\mathcal{J}(\vec{C}) = P\{\xi(y, \vec{C}) = 1\}. \tag{21.102}$$

In order to reduce this expression to the (21.63) form, additional determinate function $\Omega(\xi)$, such that

$$\Omega(\xi) = \begin{cases} 1, & \text{if } \xi = 1, \\ 0, & \text{if } \xi \neq 1. \end{cases} \tag{21.103}$$

Then, the result is

$$\mathcal{J}(\vec{C}) = M\{\Omega(\xi)\}. \tag{21.104}$$

Actually, it follows from (21.103) that

$$M\{\Omega(\xi)\} = 1P\{\xi = 1\} + 0P\{\xi = 0\} = P\{\xi = 1\},$$

which coincides with expression (21.102) for $\mathcal{J}(\vec{C})$.

So, initial expression (21.93) for a detector quality index actually /417
may be represented in the (21.104) form or, considering (21.101) and (21.103), in the form

$$\mathcal{J}(\vec{C}) = M\{\Omega(\gamma(y, \vec{C}) + \gamma_0)\}. \tag{21.105}$$

This expression with respect to its structure completely coincides with expression (21.63). Function $\Omega(\ )$ plays the role of function Q ( ) here. Since magnitude Ω may have only the values 1 and 0 [see (21.103)], then function $\Omega(\ )$ has no derivative with respect to $\vec{C}$. Therefore, as noted above, approximate relationship (21.72) should be used to determine gradient $\nabla_c \Omega(\ )$, i. e., a (21.76)-type basic algorithm:

$$\vec{C}[n] = \vec{C}[n-1] - A[n] \tilde{\nabla}_c \Omega(\gamma(y[n], \vec{C}[n-1]) + \gamma_0). \tag{21.106}$$

Here, gradient $\tilde{\nabla}_c \Omega(\gamma(y, \vec{C}) + \gamma_0)$ is determined from relationships (21.73) and (21.74),

in which Q should be replaced by Ω. In accordance with (21.70), for simplicity it is possible to assume that

$$A[n] = l\,a[n].$$

It follows from (21.106) that one must have correct (ideal) decisions γ_0 determined from relationships (21.95) to find optimum value $\vec{C}_{np}$ of vector $\vec{C}$. Evidently, it is impossible in principle to obtain such decisions in the operating mode and a special mode is required--the pattern mode. The detector may switch to the operating mode only after the pattern mode ceases, i. e., after determination of the optimum vector $\vec{C}_{np}$ value.

In light of the above, the structural diagram of a detector patterning in accordance with algorithm (21.106) may be represented in the form depicted in

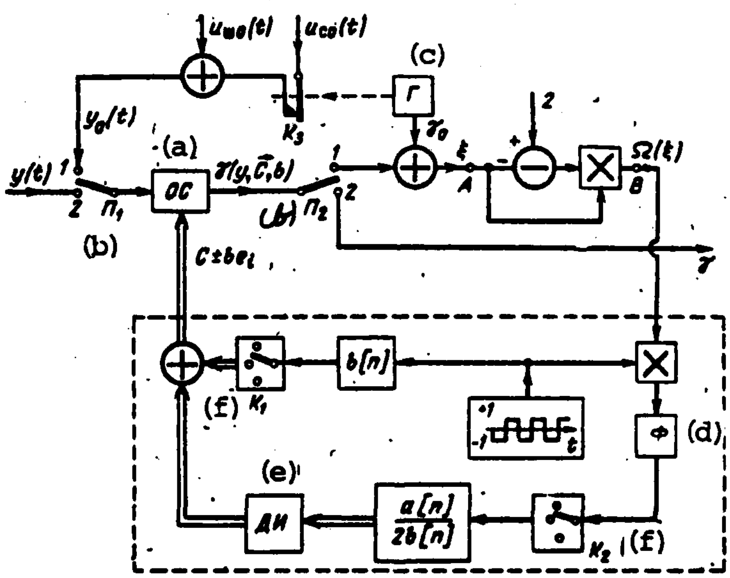


Figure 21.7. (a) -- OS [signal detector]; (b) -- P [switch]; (c) -- G [generator]; (d) -- F [filter]; (e) -- DI [discrete integrator]; (f) -- K [switch]

Figure 21.7. The circuit section enclosed by the dotted line in this figure /418 has the same purpose and composition as the Figure 21.4 circuit.

Unit *OC* -- signal detector, parameter vector $\vec{C}$ of which is optimized during the pattern process. This unit's structure is determined by decision rule (21.97) and is expanded in Figure 21.8. Here, $B_1, \ldots, B_N$ -- computers, each of which computing

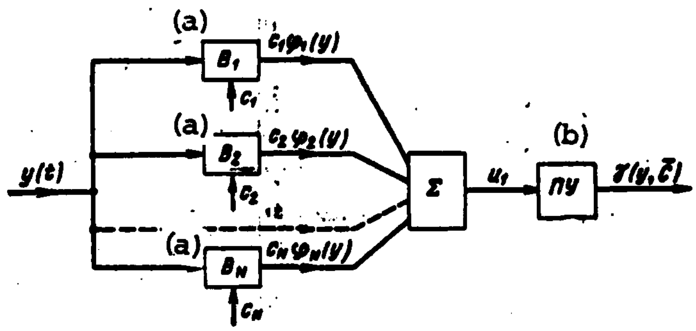


Figure 21.8. (a) -- V [computer]; (b) -- PU [threshold device].

magnitudes $\varphi_i(y)$ and multiplying the result by coefficient c_l ($l = 1, \ldots, N$). Computer output voltage sum u_1 reaches the threshold device, which supplies unity ($\gamma = 1$) when $u_1 > 0$ and zero ($\gamma = 0$) when $u_1 \leq 0$.

We now will return to the Figure 21.7 circuit. All elements in this circuit, except detector unit *OC*, may be considered an adjunct to the detector, required to accomplish detector "training." Switches Π_1 and Π_2 accomplish the switching from the pattern mode (position 1) to the operating mode (position 2).

In the pattern mode, pattern realizations $y_0(t)$ shaped from noise $u_{ш}(t)$ and signal $u_{co}(t)$ reach detector *OC* input. Signal voltage $u_{co}(t)$ is connected via switch K_3 controlled by generator Γ. This same generator processes pattern correct decisions γ_0 synchronously with the switch closing: when the switch is closed, $\gamma_0 = 1$ is issued, with $\gamma_0 = 0$ supplied in the remaining cases. The pattern comprises a series of cycles, each with duration T equalling the duration of the time that will be allocated to detection in the operating mode. In each cycle, signal $u_{co}(t)$ is either cut in or cut out, with probabilities P(C) and P(0), respectively. Therefore, generator Γ must comprise a statistical mechanism insuring random closing and opening of switch K_3 (at time T each time) in accordance with these probabilities.

The circuit portion between points A and B will process magnitude ξ of function $\Omega_{-}(\xi)$, determined by formula (21.103).

During the pattern process, detector parameters $c_1, \dots, c_N$ change with respect to time, approximating (when $t \longrightarrow \infty$) constant values $c_{\kappa p1}, \dots, c_{\kappa pN}$, which correspond to the optimum [in the sense of quality indicator (21.93)] vector $\vec{C}$ value $\vec{C}_{\kappa p}$. Essentially, patterning may be considered completed if parameters $c_1, \dots, c_N$ approximates their stationary value with sufficient accuracy (with accuracy no worse than 1%, for example). Here, Π_1 and Π_2 will convert detector *OC* to the operating mode: operating realizations y(t) are supplied to detector input, while decisions γ about signal presence ($\gamma = 1$) or absence ($\gamma = 0$) are shaped at its output.

Evidently, patterning will be effective only given signal $u_c(t)$ and noise $u_{ш}(t)$ statistical characteristics in the operating mode are identical to those in the pattern mode.

This brief examination of the stochastic approximation method makes it possible to draw the following conclusions about this method's basic advantages and disadvantages. /419

The chief advantage is that this method makes it possible relatively simply to design receiving systems under the difficult conditions of incomplete a priori information. Here, the method is applicable not only for unknown parameters of additive and non-additive noise distributions, but also for completely unknown types of these distributions. A second advantage is absence of often unsubstantiated assumptions that an unknown a priori distribution is uniform.

However, there are serious shortcomings inherent in this method (just as in any other).

First, there is the requirement to provide requisite conditions for creating the pattern mode, which is not always possible.

Second, and more important, only the parameters of a system with a known structure, rather than the system as a whole (including its structure as well),

are optimized. For example, in example 2 above, it is assumed that the detector structure has the form depicted in Figure 21.8 and, moreover, the specific form of functions (or operators) $\phi_1(y), \ldots, \phi_N(y)$ is given since only factors $c_1, \ldots, c_N$ will be subject to system design. [In example 1, this shortcoming manifests itself to a lesser degree since it is evident that, given variations, applicable for practice, in the number of parameters $c_1, \ldots, c_N$, it is possible to provide essentially any type of synthesized impulse response $\eta(t)$].

In addition, the problem of optimum selection of the discrete function a[n] type (including the magnitude of the quantization step) or of its continuous analog a(t) still is only solved partially. Extant materials mainly boil down to indications of the necessary conditions which function a[n] must meet in order to provide pattern process convergence [to (21.59)-type conditions, for instance]. But, the type of this function significantly impacts also on the pattern process duration (conversion rate) and on pattern quality overall. Therefore, it is very important that more-optimum function a[n] [or a(t)] selection methods be developed.

The problem of the noise immunity of the pattern process also has not been studied sufficiently, i. e., the problem of the impact of noise additional in comparison to that extant in the operating mode.

Finally, failure to make optimum use of extant a priori information is a very significant drawback to the stochastic approximation method. For instance, in example 2 examined above, no <u>a priori</u> (preceding initiation of patterning) information about the signal and noise is used directly when organizing the pattern process (only a priori information on signal/no signal probabilities is considered). This circumstance is an advantage of the method if such a priori information actually is completely lacking. But, this becomes a liability in those cases when a priori information is only partially, rather than completely, lacking. Evidently, /420 when even some partial a priori information is present, it must be used during selection of the designed system's <u>structure</u> (in particular, in selection of functions (or operators) $\phi_1(y), \ldots, \phi_N(y)$]. However, at present, the problem of using extant a priori information to optimize the stochastic approximation process still is far from being solved completely.

Thus, the theory of stochastic approximation still is in the initial stage of development. However, even now its use may turn out to be helpful in solving several practical problems.

21.5 Concluding Comments

The examination of various ways of designing receiving systems, given incomplete a priori information, presented in this chapter demonstrates that all these methods still are in the initial stage of development and each has signficant advantages and disadvantages. In the future, apparently, there will be succes in eliminating some of the shortcomings inherent in them. This may be achieved both through independent development of each method and by their joint supplementation and enrichment.

However, as noted in the beginning of this chapter, as a result of the infinite variety of degrees of a priori information incompleteness, there is no basis upon which to assume that, even in the future, a single system design method sufficient for <u>all</u> situations will be developed. Therefore, under complicated conditions, to increase reliability, it may be useful to undertake system design, not in accordance with one, but using at least two methods and then comparing the results obtained.

CHAPTER TWENTY-TWO /420

SPECIAL FEATURES OF RADIO RECEIVING SYSTEM DESIGN BASED ON SEVERAL QUALITY INDICES

22.1 System Design Problem Formulation Based On Several Quality Indices

Radio receiving device quality is estimated by several indices characterizing message fidelity (authenticity), operating reliability, cost, weight, overall dimensions, and so on. Message fidelity (authenticity), in turn, often must be estimated not by one, but by several indices such as false-alarm and miss /421 probability (for binary detection), root-mean-square error probability, anomalous error probability (during analog message reception), and so forth. Therefore, in general form, system design of a radio receiving device, just like any other, must be formulated in the following manner.

System operating conditions $\vec{y}$, limitations $\vec{O}$ placed on it, and quality index $\vec{K}$ are given. Here, the arrows underscore that, in the general case, there may be several conditions, limitations, and indices; for instance, vector $\vec{K}$ designates the population of all quality indices which must be considered during system design:

$$\vec{K}=\{k_1, k_2, \ldots, k_m\}. \tag{22.1}$$

Message, signal, and noise characteristics, in particular, fall in the category of conditions $\vec{y}$. Limitations $\vec{O}$ may comprise requirements for a single-channel designed system, its linearity, and so on.

The system design problem is to find that system S which, for given conditions $\vec{y}$ and limitations $\vec{O}$, provides the best quality index population $\vec{K}$ value.

Depending on the type of each elementary quality index k_i ($i = 1, \ldots, m$), its smallest or largest value is best. For instance, a system is better, the lower error probabilities $P_{лт}$ and $P_{пр}$ and the higher the probability $P_н$ of no failure (in the technological reliability sense). However, if the best value of some quality index k_i is its highest value, then it always is possible to reduce equivalent index k'_i so that its smallest value is best. For example, instead of correct detection $P_{по}$ and no failure $P_н$ probabilities, it is possible to introduce error probability $P_{ош} = 1 - P_{по}$ and no failure probability $P_0 = 1 - P_н$.

Thus, in future, it is possible without disrupting generality to assume that all quality indices $k_1, \ldots, k_m$ are positive and their <u>smallest</u> values are best.

Here, the best system would be one providing the minimum of all quality indices $k_1, \ldots, k_m$. However, in the general (nondegenerative) case, it is impossible to select a system immediately providing the minimum of two or more quality indices. Thus, no single <u>standard</u> system exists making it possible to choose the best system if a (22.1)-type <u>vector</u> quality index characterizes its quality. It is possible only to indicate several different ways to solve such a problem, each having its advantages, disadvantages, and applicability.

In future, we will use the term <u>vector</u> for system design based on population (22.1) of several quality indices, while we will use the term <u>scalar</u> for system design based on a single (one) index. In radio electronics, methods based on some sort of reduction of vector system design into scalar are the most widespread.

The first method is based on replacement of quality index population /422
(22.1) with some resultant (combined) quality index $k_р$, which is a known function

$$k_р = f(k_1, k_2, \ldots, k_m) \qquad (22.2)$$

of elementary indices $k_1, k_2, \ldots, k_m$ and design of a system based on this indicator.

The second method is based on conversion of all quality indices except one, called the main index, into a series of limitations. Then the problem will boil down to system design based on single quality index k_1, given limitations $\vec{U}'$ which includes, along with initial limitations $\vec{U}$, limitations of the following type of equalities as well

$$k_2 = k_{20};\quad k_3 = k_{30};\ \ldots;\quad k_m = k_{m0} \tag{22.3}$$

or inequalities

$$k_2 \leqslant k_{20},\quad k_3 \leqslant k_{30},\ \ldots,\quad k_m \leqslant k_{m0}, \tag{22.4}$$

or of mixed limitations.

The third method differs from the second only in that some indices k_2, . . ., k_m (or all indices) during system design are not considered at all or are considered only qualitatively (for example, during mathematical system design, system cost usually is not considered at all or is considered only to the extent that the designed system is sought in the class of single-channel or linear systems).

The first method is the most natural if the designed system (device) is part of a more-complex system characterized by quality index k_p. In this case, indices k_1, . . ., k_m may be considered parameters of the more-complex system. However, the first method has the following significant drawbacks:

1. The designed system may be so autonomous (radio measuring instrument, for example) that it cannot be considered part of a more-complex system and there is sufficient substantiation to introduce some resultant quality index k_p.

2. A more-complex system in which the given system is included often is characterized, not by one, but by several quality indices $k_{p1}, k_{p2}, \ldots, k_{pm'}$, which are dependent on indices k_1, k_2, . . ., k_m. In this case, it usually is impossible with sufficient substantiation to replace index population k_1, . . ., k_m by a single resultant index.

3. Even in those cases when the designed system is part of a more-complex

system characterized by unitary quality index k_ϕ, it may be difficult or impossible to use the first method for the following reasons:

a) if the more-complex system is developed approximately at the same time as the given system, then a (22.2)-type functional dependence usually is unknown;

b) even if the (22.2)-type dependence is known, design of the given /423 system (subsystem) based on combined criterion (22.2), being possible in principle, may turn out to be essentially very complex.

Thus, the first method, in a number of cases more natural and advisible, is far from being standard.

The second method has the following basic shortcomings:

1. In a majority of cases, it is impossible with sufficient basis to consider one quality index to be the main one and the others secondary. In addition, in the general case, it is impossible to be sure that selection of one index rather than any other as the main index will provide best system design results.

2. For quality indices $k_1, \ldots, k_m$ converted into a series of limitations, it usually is impossible unequivocally to establish their permissible values $k_{2o}, \ldots, k_{mo}$ and it is impossible to be sure that just this specific combination of these values will provide the best system design results.

The third method still to a great degree is more random than the second since it requires not only conversion of a series of quality indices into a series of limitations, but complete disregard (during the system design process) of one or several indicators.

Thus, sufficiently-substantiated reduction of vector system design into scalar system design often is impossible, especially in the initial system project planning stage. Here, it is necessary during the system design process to compare a system simultaneously with respect to all quality index populations, i. e., with respect to vector $\vec{K}$ components.

We will assume that, in an m-dimensional quality index space (Figure 22.1), one and only one vector $\vec{K}$ value corresponds to each system S and, conversely,

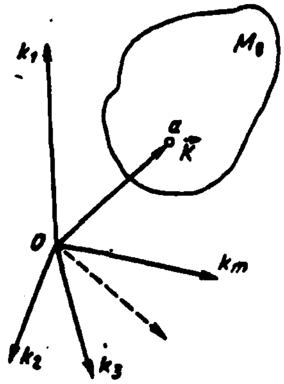


Figure 22.1

only one system S* corresponds to each point a of this space. When comparing quality indices

$$\vec{K} = \{k_1, \ldots, k_i, \ldots, k_m\},$$
$$\vec{K}' = \{k_1', \ldots, k_i', \ldots, k_m'\} \qquad (22.5)$$

we will use the following designations:

1. $\vec{K}' \leq \vec{K}$ when and only when /424

$$k_i' \leqslant k_i \quad (i = 1, \ldots, m). \qquad (22.6)$$

2. $\vec{K}' \lessdot \vec{K}$ when and only when $k_i' \leqslant k_i$ for all values of the number i, with the exception at least of one number, for which

$$k_i' < k_i. \qquad (22.7)$$

3. $\vec{K}' < \vec{K}$ when and only when

$$k_i' < k_i \quad (i = 1, \ldots, m). \qquad (22.8)$$

---

*If one and the same quality index may have an entire class of systems, then S should be understood to mean this class.

We will call system $S'(\vec{K'})$ better than system $S(\vec{K})$ when and only when even one of the system S' elementary quality indices is less then and the remaining indices not greater than the corresponding system S indices. Mathematically, this preference ratio may be written in the following form:

System $S'(\vec{K'})$ is better than $S(\vec{K})$ when and only when

$$\vec{K}' \leqslant \vec{K}. \tag{22.9}$$

This preference ratio is called unconditional since its validity is unconditional, i. e., it causes no doubts and does not require introduction of any additional conditions and assumptions for substantiation.

Evidently, the following formulation of this unconditional preference ratio completely is equivalent to the preceding formulation:

System $S'(\vec{K'})$ is worse than system $S(\vec{K})$ when and only when even one of the system S' elementary indices is greater and the rest are less than those in system $S(\vec{K})$.

In other words, system $S'(\vec{K'})$ is worse that system $S(\vec{K})$ when and only when

$$\vec{K}' > \vec{K}. \tag{22.10}$$

On the basis of preference ratio (22.9) [or (22.10)], set M_s of all possible* systems (Figure 22.1) is divided into two nonintersecting classes: set M_x of <u>worse</u> systems and set M_{xx} of <u>non-worse</u> systems. Point (system) $a \in M_s$, with quality index $\vec{K}$, by definition falls into the class of <u>worse</u> if there is even one (other) point (system) having quality index $\vec{K'} \leqslant \vec{K}$ in set M_s. Correspondingly, point (system) $a \in M_s$ falls into the <u>non-worse</u> class if no one point (system), better in comparison with it, exists in region M_s, i. e., there is no point (system) $a' \in M_s$ where $\vec{K'} \leqslant \vec{K}$.

---

*Here and in future, the term possible is used for systems satisfying given conditions $\vec{y}$ and limitations $\vec{0}$.

It follows from these determinations of worse and non-worse points (systems) that any set M_s point (system) must be either worse or non-worse, i. e., /425
belong to set M_x or M_{ux}.

Actually, we will take random point a from set M_s having some quality index $\vec{K}$. Then, evidently, only the following two mutually-exclusive situations are possible:

1. In set M_s, there is just one point a' having quality index $\vec{K'} < \vec{K}$.

2. In set M_s there is no single point a' having index $\vec{K'} < \vec{K}$.

But, if situation 1 occurs, then point a is, by definition, worse. If, conversely, situation 2 occurs, then, by definition, point a must be considered non-worse. Thus, actually any set M_s point (system) must be either worse or non-worse.

Since the object of the system design is finding the best point (system), then all worse points (systems) must be discarded and, at the next system design stage, only non-worse points considered (examined).

If we will compare among themselves any two non-worse points a$(\vec{K})$ and a'$(\vec{K'})$, then it is impossible to designate either one of them unconditionally better or worse than the other [i. e., better or worse within the bounds of preference ratio (22.9) or (22.10)]. Actually, we will assume that we recognized a'$(\vec{K'})$ as better then a$(\vec{K})$. This denotes that a point better than a exists in set M_s. Hence, it follows that point a belongs to worse points, and not to non-worse points, which contradicts the initial assertion that point a is a non-worse point.

Analogously, it also demonstrates that non-worse point a' may not be called worse than another non-worse point a. This supposition may be explained also as follows.

When comparing non-worse points a'$(\vec{K'})$ and a$(\vec{K})$, in accordance with their determination, it is mandatory that, if a lesser value of even one of the quality indices corresponds to one of the points, then a higher value of any other of

m indices will correspond to this same point. Therefore, no unconditional preference may be given to one non-worse point compared to another.

Hence, it follows that, in order to be able to select any one point (system) from the population of non-worse points (systems), some additional condition (or population of conditions) must be introduced to supplement the unconditional preference ratio. Thus, remaining within the bounds of the unconditional preference ratio, the solution to the system design problem may be reduced only to finding population (set) M_{nx} of non-worse points (systems).

In a nondegenerative case, when set M_{nx} comprises just one point, this point (system) is best (since all set M_x points are worse). Consequently, the non-worse point is best when and only when set M_{nx} is nondegenerative.

We will designate the non-worse system $\hat{S}$ and the vector quality index /426
corresponding to it

$$\hat{K} = \{\hat{k}_1, \dots, \hat{k}_m\}. \tag{22.11}$$

Then, its following property flows from the determination of the non-worse point (system) given above.

Let selected derivatives (m — 1) of fundamental indices, k_1, k_3, k_4, . . ., k_m for example, equal values $\hat{k}_1$, $\hat{k}_3$, $\hat{k}_4$, . . ., $\hat{k}_m$, respectively. Then, it is possible to assert that there is no other point (system) in set M_x which, given such index k_1, k_3, . . ., k_m values, would have the index k_2 value, which did not remain fixed identical to or less than k_2.

Actually, if <u>another</u> point would have $k_2 = \hat{k}_2$, then this would denote that all m quality indices of another point coincide with the corresponding indices of a given non-worse point. But, this would signify that "another" point in actuality coincides with the given point. If you assume that, given coinciding (m — 1) indices, another point has a lesser value of the m-th index (index k_2 in this case), then this denotes that another point is better than the given point. But, this is impossible since, based on the condition, the given point is non-worse and, consequently, no better point than it exists.

For example, consider that m = 3 and we found non-worse point $a(\hat{k}_1, \hat{k}_2, \hat{k}_3)$ (Figure 2.22). Then, the following assertions are valid:

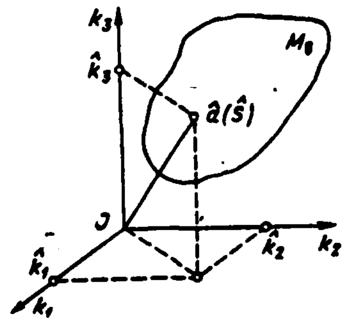


Figure 22.2

1. For the givens $(\hat{k}_1, \hat{k}_2)$, no point (system) which would have an index k_3 value less than or equal to $\hat{k}_3$ exists in space M_s. In other words, the ordinate of point $\hat{a}$ indicates the minimum-possible index k_3 value which may be achieved for given conditions $\vec{y}$, limitations $\vec{O}$, and fixed values (k_1, k_2) of the remaining quality indices.

2. For givens $\hat{k}_2$, $\hat{k}_3$, the minimum-possible index k_1 value equals $\hat{k}_1$.

3. For the givens $(\hat{k}_3, \hat{k}_1)$, the minimum-possible index k_2 value equals $\hat{k}_2$.

Evidently, in the case of m quality indices, it is possible to make m analogous assertions.

It follows from the aforementioned non-worse point property that a minimum-possible (potential) value of any of the quality indices, given fixed values for all remaining (m — 1) indices, corresponds to each such point. Consequently, if one succeeds in finding set $M_{\text{нх}}$ of non-worse points (systems), then this /427 denotes that one may succed in finding m dependences of the type

$$\left.\begin{array}{l} k_{1\,\text{мин}} = f_1(k_2, k_3, \ldots, k_m), \\ k_{2\,\text{мин}} = f_2(k_1, k_3, \ldots, k_m), \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ k_{m\,\text{мин}} = f_m(k_1, k_2, \ldots, k_{m-1}), \end{array}\right\} \tag{22.12}$$

where values $k_{1\,нхн}, k_{2\,нхн}, \ldots, k_{m\,нхн}$ may be called particular potential values of quality indices $k_1, k_2, \ldots, k_m$ (as opposed to a single, global, potential value which will be found in a case of system design with respect to a single index).

The following basic properties of non-worse points may be formulated on the basis of what has been presented:

Property 1. All set M_s points (systems) that are non-worse are worse (in the sense of vector quality index $\vec{K}$).

Property 2. No non-worse point (system) may be recognized as better or worse than another non-worse point (system).

Property 3. If only one non-worse point (system) exists, then it is the best.

Property 4. A minimum (best-possible, potential) value of any quality index k_i ($i = 1, \ldots, m$) corresponds to each non-worse point (system), given fixed values of all remaining ($m - 1$) quality indices.

Several additional important properties of non-worse points, methods of finding these points, and several examples are examined in the following sections. Here, further examination is limited to cases of two and three quality indices in accordance with the following considerations:

1. Cases of mathematical system design with respect to two or three quality indices are encountered most often in practical problems.

2. When more than three quality indices are involved, system design results using the unconditional preference method are considerably less clear and, in particular, do not allow clear geometric interpretation.

22.2 Two Quality Indices

In the case of two quality indices ($m = 2$), general expressions (22.12) take

the following form:

$$k_{1\,\text{мин}} = f_1(k_2), \tag{22.13a}$$

$$k_{2\,\text{мин}} = f_2(k_1). \tag{22.13b}$$

The set of non-worse points must satisfy both these relationships, i. e., must possess the following properties:

1. For each index k_2 value, value k_1 must be the minimum-possible /428
value, i. e, for a given k_2, no points (systems) less than k_1 must exist.

2. For each index k_1 value, value k_2 must be minimum, i. e., for a given k_1, no points less than k_2 must exist.

It is easy to demonstrate that non-worse points may be those and only those points which, on quality index plane $k_1\ 0\ k_2$ (Figure 22.3), are located on a monotone

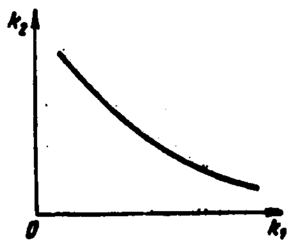


Figure 22.3

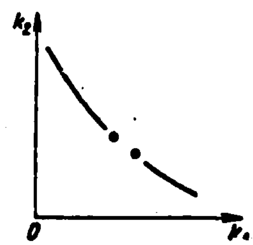


Figure 22.4

nonincreasing curve, both when k_1 rises and when k_2 rises.

Consequently, in a case of two quality indices, the set of non-worse points has the form of a monotone nonincreasing curve, both when k_1 rises and when k_2 rises. It is not mandatory that this curve be continuous (for example, see Figure 22.4) and, in several cases, it may comprise only several points (Figure 22.5). However, it always must meet the aforementioned condition of strict monotonicity.

The only exception is a nondegenerative case in which set $M_{нх}$ comprises just one point. In future, we will restrict ourselves to examination only of those cases when a set of non-worse points forms a continuous curve.

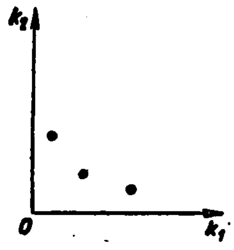


Figure 22.5

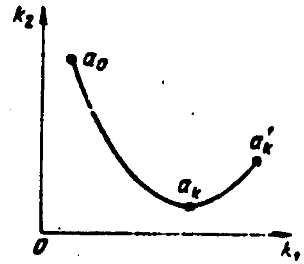


Figure 22.6

It follows also from determination of non-worse points (systems) that, for given conditions $\vec{y}$ and limitations $\vec{0}$, no single point (system) located to the left or below the given curve exists (the validity of this assertion is demonstrated easily from the converse). Therefore, set $M_{нх}$ of non-worse points (systems) is the <u>lower left boundary</u> of region $M_в$ of all possible points (systems).

Thus, in a case of two quality indices, it is possible to add the following properties to the properties of non-worse points indicated on page 613:

<u>Property 5</u>. Set $M_{нх}$ of non-worse points forms the lower left boundary of region $M_в$ of all possible points.

<u>Property 6</u>. The lower left boundary is a monotone nonincreasing curve, both when k_1 rises and when k_2 rises, i. e., it has the property of strict monotonicity.

It follows from the aforementioned that system design with respect to /429 two quality indices in essence will boil down to finding the lower left boundary providing optimum (best-possible) link

$$k_1 = f_{опт}(k_2) \tag{22.14}$$

between both quality indices. There are different methods for finding the lower left boundary. We will examine only the two most-widespread approaches.

The <u>first method</u> is that the minimum of one of the quality indices is sought, given a fixed, but random, value of the other:

$$k_1 = \min, \text{ where } k_2 = \text{const}, \tag{22.15a}$$

or

$$k_2 = \min, \text{ where } k_1 = \text{const}. \tag{22.15b}$$

Such minimization is accomplished for all possible (or of interest) values of the fixed (i. e., converted into a series of limitations) quality index. Here, if indicator k_2 is minimized, then the result is a dependence of the type

$$k_{2\,\text{мин}} = f_{p1}(k_1). \tag{22.16a}$$

If k_1 is minimized, then the result is

$$k_{1\,\text{мин}} = f_{p2}(k_2). \tag{22.16b}$$

We will call dependences (22.16a) and (22.16b) type I and II performance curves, respectively.

It follows from comparison of expressions (22.13a), (22.13b), and 22.16a) that a type I performance curve satisfies one of the two conditions placed on non-worse points (condition (22.13b), but may not satisfy the second (condition 22.13a). The following important type I performance curve property emerges:

The performance curve will comprise all non-worse points, but may comprise several others, i. e., worse points, in addition.

For example, performance curve $a_o\ a_k\ a_k'$ depicted in Figure 22.6 will comprise section $a_k\ a_k'$ consisting of a population of worse points in addition to lower left boundary $a_o\ a_k$.

A type II performance curve also has an analogous property. /430

Consequently, each of the performance curves (22.16a) and (22.16b) will comprise all lower left boundary points and, in addition, in several cases also may comprise a series of additional, worse, points. These additional points are easy to eliminate since they are located in sectors where the monotone nonincreasing nature of the k_1 dependence on k_2 (or k_2 on k_1) is disrupted. It may be demonstrated that the performance curve completely coincides with the lower left boundary when and only when this performance curve has a monotone nonincreasing nature in the entire range of possible (or of interest to us) quality index values.

These special performance curve features make it possible to draw the following conclusions important in practice:

1. Conversion of any of the quality indices (k_1 or k_2) into a series of limitations does not hinder obtaining optimum system design results if, during such a conversion, the minimum of one of the indices (k_1, for instance) is sought for all possible (or of interest to us) fixed values of the other index (k_2, for example).

2. Elimination of additional systems (worse points) in any of the performance curves suffices for finding the lower left boundary, i. e., the population of non-worse systems.

3. If the performance curve is a monotone nonincreasing function (in the entire range of the values of its argument of interest to us), then it completely coincides with the lower left boundary and, consequently, there is no requirement for elimination.

4. Each lower left boundary point has the six important properties indicated on pages 612 and 614, particularly property 4, which is very important in practice.

5. If a performance curve [characteristic $k_{2\,\text{мин}} = f_{\text{p1}}(k_1)$, for example] is a monotone nonincreasing function, then it provides best results, not only in the sense of that quality index was minimized (index k_2, for instance), but also in the sense of the second quality index (converted into a series of limitations).

It should be noted that, in practical problems, a performance curve has a monotone nonincreasing nature in a majority of cases. This is because an increase in one quality index denotes a decrease in requirements levied on the system relative to this index and, therefore, makes it possible, as a rule, to realize the best (lesser) value of another index. Only in relatively-rare cases may the population of limitations $\vec{O}$ be such that its strict monotonicity is disrupted in individual performance curve sections. Therefore, one may consider the monotone nonincreasing nature of a performance curve a rule that is disrupted only in rare cases.

We will examine as our illustration use of the method examined, which we will call the performance curve method, for design of a binary signal detector system in accordance with the minimum composite error probability $P_{\text{ош}}$ criterion. /431 This, in essence, was the system design problem formulation presented in preceding chapters (Chapters 5, 9, and 14):

Find the structure of a detector providing minimum composite probability $P_{\text{ош}}$ of detection error for a given (fixed) signal-to-noise ratio q value.

The detector is this problem is characterized by two quality indices, $P_{\text{ош}}$ and q; the lower each of these indices, the better the detector. Consequently, in this case, it is possible to assume

$$k_1 = P_{\text{ош}}, \quad k_2 = q. \tag{22.17}$$

System design is minimization of one of these indices (k_1) for a fixed, but random, value of the second (k_2), i. e., for the performance curve method presented above. As a result of the system design, a type II performance curve was obtained:

$$k_{1\,\text{мин}} = f_{p2}(k_2).$$

When the signal fluctuates slowly and P(O) = P(C) = 0.5, it has the form depicted in Figure 9.8 (here, $q = Q_{\text{ср}}/N_0$). Since this characteristic has a monotone nonincreasing nature, then, by virtue of the aforementioned, it coincides with the lower left boundary and, consequently, is the geometric location of all non-worse points (systems). In this case, all non-worse systems (points) differ only in threshold U_0 magnitude.

Since the Figure 9.8 characteristic coincides with the lower left boundary, it has all its inherent properties (see pages 612 and 614) and, in particular, this property important in practice: for every q value, it provides the minimum (with respect to all possible detectors) magnitude of composite error probability P_{om} and, for each given value of magnitude P_{om} --the minimum (with respect to all possible detectors) value of the requisite signal-to-noise ratio q. In other words, if we would design a detector system, not in accordance with criterion P_{om} = min (where q = const), but in accordance with the criterion q = min (where P_{om} = const) or we attempted to consider both quality indices (P_{om} and q) immediately during the design process, then we would not be able to obtain a detector with the best ratio between P_{om} and q, with a lesser q value where P_{om} = const in particular.

We now will examine a second method of finding the lower left boundary. It involves finding the population of systems providing minimum quality index linear combination

$$k_n = k_1 + ck_2 \tag{22.18}$$

given all possible positive finite values of weight factor c fixed during the minimization process, i. e., given

$$0 < c < \infty. \tag{22.19}$$

We will use k_{1c} and k_{2c} to designate the values of quality indices k_1 and k_2 corresponding to that system S_c, which provides the magnitude k_n minimum for a given weight factor c value. Varying value c within the (22.19) range, we obtain dependences of the type

$$k_{1c} = f_1(c), \tag{22.20a}$$

$$k_{2c} = f_2(c). \tag{22.20b}$$

Eliminating magnitude c from these relationships, we will obtain

$$k_{1c} = f_c(k_{2c}). \tag{22.21}$$

We will call this dependence the weight characteristic. As analysis demonstrates, a weight characteristic has the following properties:

1. It either completely coincides with the lower left boundary or is /432 a part of it.

2. A sufficient condition for the coincidence* of the weight characteristic with the lower left boundary is coherence of the characteristic (in the entire range of possible quality indices or of those of interest to us).

Here and in future, the weight characteristic is called linked if it is determined on a linked set of points. (A point set is called coherent if it cannot be represented in the form of a union of two disjoint sets, each of which not comprising threshold points of the other).

3. A condition required for coincidence* of the weight characteristic with the lower left boundary is convexity** of the lower left boundary.

Important consequences flow from these properties:

1. A weight characteristic may not comprise all non-worse points (systems), but it must have one worse point (system).

2. If a weight characteristic is coherent (in the entire range of possible quality index values or of those of interest to us), then it will comprise all non-worse points (systems) rather than one worse point (system), i. e., it completely will solve the problem of system design based on two quality indices.

As analysis shows, the sufficient condition of coincidence of the weight characteristic with the lower left boundary is met in a majority of practical problems (but not in all problems). Therefore, a system design approach based on minimization of the sum (22.18) usually is effective.

---

*Precise to extreme points usually noexistent in practice.

**The line (curve, function) specified in segment x is called convex (downward) if all internal points of any arc of its graph (in region x) are located beneath the chord linking the ends of the arc or on this arc.

We will examine the following case as our illustration.

Consider design of a binary detector system for a slowly-fluctuating signal on a background of white noise based on two quality indices

$$k_1 = P_{\text{пр}} \text{ and } k_2 = P_{\text{лт}}, \tag{22.22}$$

where $P_{\text{пр}}$ and $P_{\text{лт}}$ -- conditional signal miss and false-alarm probabilities (signal-to-noise ratio q at detector input is assumed to be given). A priori probabilities of signal presence and absence and detection error weight are unknown so there is no way to form a single combined quality index.

The requirement is to find the population of non-worse systems and the potential (best-possible) relationship between quality indices k_1 and k_2 corresponding to this population.

We will solve this problem through minimization of expression (22.18), which in this case has the form

$$k_{\Sigma} = P_{\text{пр}} + cP_{\text{лт}}. \tag{22.23}$$

Solution of this problem provides the following results (see Chapters 9 and 14).

The optimum system will comprise matched linear filter, inertialess /433
envelope detector, and output threshold bias. Given a linear detector with a unitary transfer constant, threshold magnitude equals

$$U_{0c} = \sqrt{2}U_{\text{ш}}\sqrt{\left(1+\frac{1}{q}\right)\ln\left[c\,(1+q)\right]} \tag{22.24}$$

(one should assume $U_{oc} = 0$ where $c < \frac{1}{1+q}$) and is the single system parameter dependent on weight factor c.

Error probabilities, for the given weight factor c value, equal

$$P_{\text{пр}\,c}=1-e^{-\frac{U_{0c}^2}{2U_{\text{ш}}^2(1+q)}},\qquad P_{\text{лт}\,c}=e^{-\frac{U_{0c}^2}{2U_{\text{ш}}^2}}, \tag{22.25}$$

where q -- signal-to-noise power ratio at system input; $U_{\text{ш}}$ -- envelope detector output noise voltage mean-square value.

It follows from (22.25)

$$P_{\text{пр}\,c}=1-P_{\text{лт}\,c}^{\frac{1}{1+q}}. \tag{22.26}$$

It is easy to become convinced that this function is linked in the entire range of quality index values of interest to us $(P_{\text{пр}}<1,\ P_{\text{лт}}<1)$. Consequently, it completely coincides with the lower left boundary for any $q > 0$. (Where $q = 0$, the examination loses meaning, both from a physical and a mathematical point of view since, here, in accordance with (22.24), the result is $U_{oc} \to \infty$, i. e., the parameter of the non-worse system determining its characteristic will become indeterminate).

It also is easy to become convinced that the lower left boundary found in this case actually is strictly monotonic--magnitude $P_{\text{пр}\,c}$ monotonically decreases when probability $P_{\text{лт}\,c}$ rises.

Thus, in this example of binary detector system design, use of the method of weighted sum minimization (22.18) made it possible to find the population of non-worse systems. In this case, one non-worse system differs from another only in the threshold bias U_{oc} value, which may be determined from formula (22.4) or, considering (22.25), from the formula

$$U_{0c}=U_{\text{ш}}\sqrt{2\ln\frac{1}{P_{\text{лт}\,c}}}. \tag{22.27}$$

The link between the quality indices of the populationof non-worse systems found is expressed by relationship (22.26). Here, in accordance with the aforementioned non-worse system property 4, the minimum-possible value of magnitude $P_{\text{лт}}$ for given magnitude $P_{\text{пр}}$ and the minimum value of probability $P_{\text{пр}}$ for the given probability $P_{\text{лт}}$ value corresponds to each such system (i. e., to each threshold U_{oc} value, in particular).

In conclusion, we will note the following comparative special features of the methods examined for finding non-worse systems.

Methods based on finding the performance curve [relationship (22.16)] /434 in the general case provide the same set of systems which will comprise all non-worse systems, and, in addition, there may be several worse systems. The method based on minimization of weighted sum (22.18), on the other hand, may (but this is not mandatory) lead to a loss of part of the non-worse systems, but does not provide a single worse system.

The drawback of the latter is more significant since worse systems usually may be discovered without any special effort and screened out during subsequent system design; missing some of the non-worse systems may not be counterbalanced and system design results may deteriorate. Therefore, in those cases when use of the first method provides success in minimization without special difficulties, its use is preferable. However, when meeting several of the aforementioned additional conditions [strict performance curve monotonicity or characteristic $k_{1c} = f_c(k_{2c})$ coherence], both methods make it possible to find the population of non-worse systems with the danger of losing part of the non-worse systems or contaminating the population of worse systems found.

22.3 System Design Based On Three Quality Indices

In a case of system design based on three quality indices, general relationships (22.12) for non-worse systems take the following form:

$$\left.\begin{aligned} k_{1\,\text{мин}} &= f_1(k_2, k_3), \\ k_{2\,\text{мин}} &= f_2(k_3, k_1), \\ k_{3\,\text{мин}} &= f_3(k_1, k_2). \end{aligned}\right\} \qquad (22.28)$$

The set of non-worse systems meeting these conditions in three-dimensional quality index space form some surface

$$k_1 = f_{\text{опт}}(k_2, k_3), \tag{22.29}$$

which we will call the optimum surface (this surface is depicted by the hatched

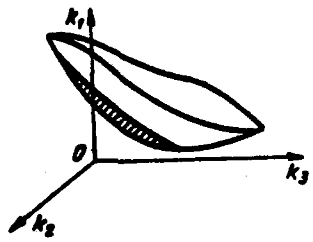


Figure 22.7

space in Figure 22.7). Evidently, this surface is a generalization of the concept of the lower left boundary a case of three quality indices. Each surface point has the four properties indicated in § 22.1.

We will limit ourselves to examination of cases in which the optimum surface is continuous in the entire range of possible (or of interest to us) quality index values. Then, considering the properties of non-worse systems, it is possible to become convinced that no single point (system) from set M_s of all possible systems exists between the origin and the optimum surface. Therefore, the optimum surface will belong to the external set M_s boundary.

We will introduce the concept of an operating surface for further examination. The operating surface is that population of points possessing the minimum value of one quality index, given fixed (but random) values for the other indices in the entire range of possible (or of interest to us) quality index values. /435

Then, depending on which of the three indices is subject to minimization (as the operating surface is formed), we will have the following operating surfaces:

a) Primary operating surface

$$k_{1\,\text{мин}} = f_{1p}(k_2, k_3); \tag{22.30a}$$

b) Secondary operating surface

$$k_{2\,\text{мин}} = f_{2p}(k_3, k_1); \tag{22.30b}$$

c) Tertiary operating surface

$$k_{3\,\text{мин}} = f_{3p}(k_1, k_2). \tag{22.30c}$$

It follows from comparison of expressions (22.28) and (22.30) that each operating surface will comprise all points of the optimum surface (i. e., all non-worse points) and, in addition, may comprise several additional (worse) points.

Hence, in particular, the following conclusion important for practice emerges: if during system design one provides minimization only based on one quality index, given fixed (but random) values for the other two indices, then this will not lead to a loss of any of the non-worse points (systems).

In view of the symmetry of relationships (22.30a)--(22.30c), it is possible in future, without a loss in generality, to restrict ourselves to examination only of a primary operating surface. The equation for this surface may be represented on a plane in the form of either of the following two families:

a) $$k_{1\,\text{мин}} = f_{1p}(k_2, k_{3m}); \tag{22.31a}$$

b) $$k_{1\,\text{мин}} = f_{1p}(k_{2m}, k_3). \tag{22.31b}$$

Here, the index «m» denotes which of the two function (22.30a) arguments plays the role of the parameter of the family of characteristics (Figure 22.8).

We will call these families type I and II performance curve families, respectively. It may be demonstrated that performance curve families have the following properties:

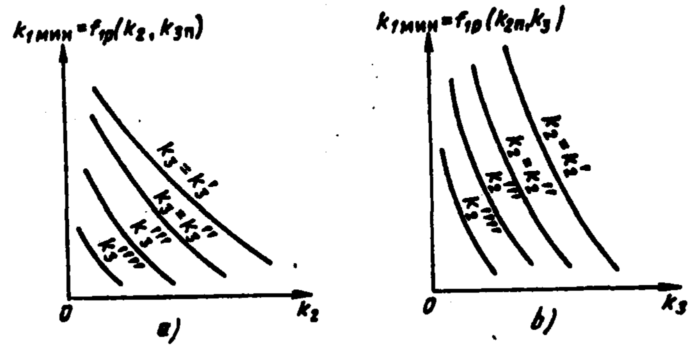


Figure 22.8

1. Each family will comprise an entire set $M_{нх}$ of non-worse points (systems) and, in addition, may comprise several additional (worse) points (systems).

2. A necessary and sufficient condition of complete operating surface (performance curve family) coincidence with the optimum surface may be formulated in the following way: all type I and II performance curve families [(22.31b)] must be monotone nonincreasing functions in the entire range of possible (of interest to us) quality index values. In practical problems, the condition is met as a rule (but not always). /436

The following consequences important for practice stem from these properties:

1. It is possible when seeking the population of non-worse systems (the optimum surface) to use the determination (computation) of the primary (secondary, tertiary) operating surface, i. e., to find a (22.30a)-type dependence. This operating surface (and the type I and II performance curve family corresponding to it) will comprise all the non-worse points (systems) and, in addition, may comprise several worse points (systems), which must be screened out.

2. If the type I and II performance curves corresponding to the operating surface are monotone nonincreasing curves in the entire range of possible (of interest to us) quality index values, then they will not comprise worse points and, consequently, no screening is required.

3. When meeting the strict monotonicity condition in the preceding paragraph,

each point (system) of the performance curve family is non-worse and has all the aforementioned properties of non-worse points (systems), the following property in particular:

a) for the given index k_2 and k_3 values, it provides the minimum-possible (among all possible systems) index k_1 value;

b) the minimum k_2 is supplied for the given k_3 and k_1;

c) the minimum k_3 is supplied for the given k_1 and k_2.

We will examine as our example design of a binary detector system for unknown a priori signal presence and absence probabilities.

In this case, mathematical system design requires consideration of three quality indices: signal miss probability $P_{\text{пр}}$, false-alarm probability /437 $P_{\text{лт}}$, and requisite signal-to-noise ratio q. The less the value of these indices, the better the system. Therefore, it is possible to designate

$$k_1 = P_{\text{пр}}, \quad k_2 = P_{\text{лт}}, \quad k_3 = q. \tag{22.32}$$

It follows from what has been stated that it is possible during the system design, without danger of losing optimum decisions (non-worse systems), to minimize any of these indices, assuming that the rest are fixed (but random) magnitudes.

When using the Neyman-Pearson criterion examined in Chapter 14, magnitude $P_{\text{пр}}$ is minimized for fixed values of $P_{\text{лт}}$ and q, i. e., the operating surface equation will be found

$$k_{1\,\text{мин}} = f_{1\text{p}}(k_2, k_3) \tag{22.33}$$

and the performance curve (22.31a) and (22.31b) families corresponding to it. The type I and II performance curve family of the first type qualitatively is depicted in Figure 14.5, while it is depicted quantitatively in Figure 9.6 (for a random-phase signal). However, magnitude (1-- k_1), rather than k_1, is plotted on the X-axis in these figures. The family of Figure 9.6 performance curves plotted in coordinates (k_1, k_2) has the form depicted in Figure 22.9a.

Based on this figure, it is not difficult to plot the corresponding type II

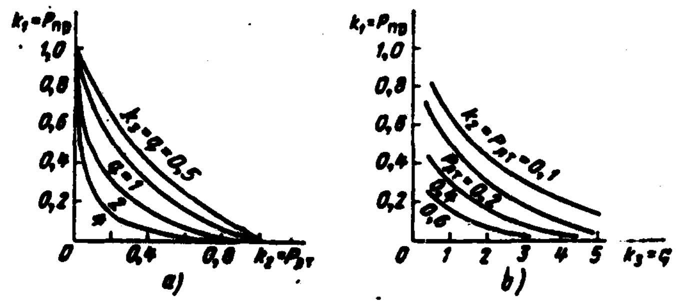


Figure 22.9

performance curve family depicted in Figure 22.9b. As is evident, any curve of these families has a monotone nonincreasing nature. Here, all non-worse points (systems) differ only in threshold bias U_0 magnitude.

The performance curves depicted in Figure 22.9 correspond to a signal with known amplitude and random phase. For a fluctuating signal where $P_{np} < 0.1$, the operating surface is determined from relationship (9.83), from which it follows that (where $P_{np} < 0.1$)

$$P_{np} \approx \frac{\ln(1/P_{лт})}{q},$$

where $q = Q_{cp}/N_0$.

In this case, it is evident even without graphic representation that magnitude P_{np} is monotone nonincreasing, both when $P_{лт}$ rises (where q = const) and when q rises (where $P_{лт}$ = const), i. e., type I and II performance curves have a monotone nonincreasing nature.

Thus, the necessary and sufficient condition of complete operating surface coincidence with the optimum surface is met in all the cases examined. This signifies that a family of performance curves of any type (i. e., types I and II) will comprise all possible non-worse points (systems) and will not comprise a single worse point. Hence, the following important conclusion stems from consideration of property 4 of non-worse systems.

A detector system designed in accordance with the Neyman-Pearson criterion /438 provides, not only the minimum magnitude $P_{\text{пр}}$ value for given q and $P_{\text{лт}}$, but the minimum $P_{\text{лт}}$ value for given q and $P_{\text{пр}}$ and the minimum q value for given $P_{\text{лт}}$ and $P_{\text{пр}}$. Moreover, it follows from the aforementioned that, in spite of the fact that system design is accomplished through minimization only of one of three quality indices (magnitude $P_{\text{пр}}$) when the Neyman-Pearson criterion is used, results obtained may not be improved if all three quality indices ($P_{\text{лт}}$, $P_{\text{пр}}$, and q) are considered together during system design (selection of the best system).

We examined the method of finding the optimum surface (non-worse system population) by finding the operating surface equation. However, as was true in the case of two quality indices, there are other approaches to finding the optimum surface. In particular, it is possible when meeting the corresponding convexity conditions to use the quality index linear sum minimization method

$$k_{\text{л}} = k_1 + c_1 k_2 + c_2 k_3,$$

where weights c_1 and c_2 range from $0 < c_{1,2} < \infty$.

In conclusion, we must note that the theory of radio system design based on several quality indices is only in the initial stage of development and the material presented in this chapter should be looked upon only as an introduction to this theory.

APPENDIX

$\chi^2_{2m}(\alpha)$ TABLE OF DISTRIBUTION

$2m$ \ α	0.99	0.98	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41	3,22	4,60	5,99	7,82	9,21	11,6	12,4	13,8
4	0,297	0,429	0,711	1,064	1,649	2,20	3,36	4,88	5,99	7,78	9,49	11,67	13,28	14,9	16,9	18,5
6	0,872	1,134	1,635	2,20	3,07	3,83	5,35	7,23	8,56	10,64	12,59	15,03	16,81	18,6	20,7	22,5
8	1,646	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,03	13,36	15,51	18,17	20,1	21,9	24,3	26,1
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,78	13,44	15,99	18,31	21,2	23,2	25,2	27,7	29,6
12	3,57	4,18	5,23	6,30	7,81	9,03	11,34	14,01	15,81	18,55	21,0	24,1	26,2	28,3	31	32,9
14	4,66	5,37	6,57	7,79	9,47	10,82	13,34	16,22	18,15	21,1	23,7	26,9	29,1	31	34	36,1
16	5,81	6,61	7,96	9,31	11,15	12,62	15,34	18,42	20,5	23,5	26,3	29,6	32,0	34	37	39,2
18	7,02	7,91	9,39	10,86	12,86	14,44	17,34	20,6	22,8	26,0	28,9	32,3	34,8	37	40	42,3
20	8,26	9,24	10,85	12,44	14,58	16,27	19,34	22,8	25,0	28,4	31,4	35,0	37,6	40	43	45,3
22	9,54	10,60	12,34	14,04	16,31	18,10	21,3	24,9	27,3	30,8	33,9	37,7	40,3	42,5	46	48,3
24	10,86	11,99	13,85	15,66	18,06	19,94	23,3	27,1	29,6	33,2	36,4	40,3	43,0	45,5	48,5	51,2
26	12,20	13,41	15,38	17,29	19,82	21,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6	48	51,5	54,1
28	13,56	14,85	16,93	18,94	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3	51	54,5	56,9
30	14,95	16,31	18,49	20,6	23,4	25,5	29,3	33,5	36,2	40,3	43,8	48,0	50,9	54	57,5	59,7

BIBLIOGRAPHY

1. Котельников В. А. Теория потенциальной помехоустойчивости. Госэнергоиздат, 1956. /440

2. Вудворд Ф. М. Теория вероятностей и теория информации с применениями в радиолокации. Пер. с англ. Изд-во «Советское радио», 1955.

3. Вайнштейн Л. А. и Зубков В. Д. Выделение сигналов на фоне случайных помех. Изд-во «Советское радио», 1960.

4. Харкевич А. А. Очерки общей теории связи. Гостехиздат, 1955.

5. Фалькович С. Е. Прием радиолокационных сигналов на фоне флуктуационных помех. Изд-во «Советское радио», 1961.

6. Долуханов М. П. Введение в теорию передачи информации по электрическим каналам связи. Связьиздат, 1955.

7. Смирнов В. А. Основы радиосвязи на УКВ. Связьиздат, 1957.

8. Гольдман С. Теория информации. Пер. с англ. Изд-во иностранной литературы, 1957.

9. «Пороговые сигналы». Пер. с англ. Изд-во «Советское радио», 1952.

10. Цянь Сюэ-сэнь. Техническая кибернетика. Пер. с англ. Изд-во иностранной литературы, 1956.

11. Пугачев В. С. Теория случайных функций и ее применение к задачам автоматического регулирования. Гостехиздат, 1957.

12. Солодовников В. В. Введение в статистическую динамику систем авторегулирования. Гостехиздат, 1952.

13. Перов В. П. Статистический синтез импульсных систем. Изд-во «Советское радио», 1959.

14. Блекуэлл Д. и Гиршик М. А. Теория игр и статистических решений. Пер. с англ. Изд-во иностранной литературы, 1958.

15. «Сборник трудов научно-технического об-ва радиотехники и электросвязи им. А. С. Попова». Под ред. В. И. Сифорова. Вып. II, III. Госэнергоиздат, 1958, 1959.

16. «Прием сигналов при наличии шума». Сборник переводов под ред. Л. С. Гуткина. Изд-во иностранной литературы, 1960.

17. «Теория информации и ее приложения». Сборник переводов под ред. А. А. Харкевича. Физматгиз, 1959.

18. «Теория передачи электрических сигналов при наличии помех». Сборник переводов под ред. Н. А. Железнова. Изд-во иностранной литературы, 1953.

19. «Прием импульсных сигналов в присутствии шумов». Сборник переводов под ред. А. Е. Башаринова и М. С. Александрова. Госэнергоиздат, 1960.

20. Бунимович В. И. Флуктуационные процессы в радиоприемных устройствах. Изд-во «Советское радио», 1951.

21. Левин Б. Р. Теория случайных процессов и ее применение в радиотехнике Изд-во «Советское радио», 1960.

22. Лебедев В. Л. Случайные процессы в электрических и механических системах. Физматгиз. 1958.

23. Крамер Г. Математические методы статистики. Пер. с англ. Изд-во иностранной литературы, 1948.

24. Смирнов Н. В. и Дунин-Барковский И. В. Краткий курс математической статистики для технических приложений. Физматгиз, 1959.

25. Гнеденко Б. В. Курс теории вероятностей. Гостехиздат, 1950.

26. Вентцель Е. С. Теория вероятностей. Физматгиз, 1958.

27. Рыжик И. М. и Градштейн Н. О. Таблицы интегралов, сумм, рядов и произведений. Гостехиздат, 1951.

28. Wiener N. Extrapolation, interpolation and smoothing of stationary time series. John Wiley, New-York, 1949.

29. Вальд А. Последовательный анализ. Физматгиз, 1960.

30. Wald A. Statistical decision functions. John Wiley, New-York, 1947.

31. Миддлтон Д. Введение в статистическую теорию связи. Т. I и II. Изд-во «Советское радио», 1961, 1962.

32. Давенпорт В. Б., Рут В. Л. Введение в теорию случайных сигналов и шума. Изд-во иностранной литературы, 1960.

33. Котельников В. А. О пропускной способности «эфира» и проволоки в электросвязи. Всесоюзн. энерг. ком., Материалы к первому всесоюзному съезду по вопр. реконстр. дела связи. Изд. Упр. связи РККА, 1933.

34. Колмогоров А. Н. Интерполирование и экстраполирование стационарных случайных последовательностей. Известия АН СССР, сер. математическая, 1941, т. 5, № 1.

35. Сифоров В. И. О влиянии помех на прием импульсных радиосигналов. «Радиотехника», 1946, № 1.

36. Белоусов А. П. О наивысшей реальной чувствительности приемника. «Радиотехника», 1946, № 5.

37. Котельников В. А. Проблемы помехоустойчивости радиосвязи. Радиотехнический сборник. Госэнергоиздат, 1947.

38. Харкевич А. А. Обнаружение слабых сигналов. «Радиотехника», 1953, № 5.

39. Карновский М. И. О подавлении флуктуационных помех при корреляционном методе приема. «Радиотехника», 1954, № 3.

40. Вощкий В. С. Обнаружение слабых сигналов способом асинхронного накопления. «Радиотехника», 1954, № 6.

41. Филиппов Л. И. О помехоустойчивости импульсного радиоприема. «Радиотехника», 1954, № 6.

42. Уркович Г. Фильтры для обнаружения слабых радиолокационных сигналов на фоне мешающих отражений. «Вопросы радиолокационной техники», 1954, № 2.

43. Филиппов Л. И. Потенциальная помехоустойчивость при приеме импульсных радиосигналов. «Радиотехника», 1955, № 10.

44. Шастова Г. А. Исследование помехоустойчивости передачи команд телеуправления методами теории потенциальной помехоустойчивости. «Автоматика и телемеханика», 1955, № 4.

45. Чайковский В. И. Прием импульсных сигналов по методу взаимной корреляции. «Радиотехника», 1955, № 6.

46. Веннер А. и Древник Р. Проблема оптимального обнаружения импульсных сигналов в шуме. «Вопросы радиолокационной техники», 1956, № 5.

47. Чайковский В. И. Помехоустойчивость фильтрового автокорреляционного приемника импульсных сигналов. «Радиотехника», 1956, № 4.

48. Башаринов А. Е. О помехоустойчивости корреляционного метода приема. «Радиотехника», 1956, № 5.

49. Зюко А. Г. Помехоустойчивость и эффективность фототелеграфной радиосвязи при флуктуационных помехах. «Радиотехника», 1956, № 8.

50. Зюко А. Г. Автореферат о расчете помехоустойчивости радиоприемника при большом уровне флуктуационных помех. «Радиотехника», 1956, № 10.

51. Харкевич А. А. О вычислении спектров случайных процессов. «Радиотехника», 1957, № 5.

52. Железнов Н. А. О принципиальных вопросах теории сигналов и задачах ее дальнейшего развития на основе новой стохастической модели. «Радиотехника», 1957, № 11.

53. Карновский М. И. и Чайковский В. И. Метод повышения помехоустойчивости автокорреляционного приема импульсных сигналов. «Радиотехника», 1957, № 2.

54. Сверлинг П. Максимальная точность определения координат /442
импульсной радиолокационной станции. «Вопросы радиолокационной техники», 1957, № 2.

55. Теплов Н. Л. К оценке помехоустойчивости методов радиоприема, основанных на усреднении функций сигнала и помехи. «Радиотехника», 1957, № 9.

56. Финкельштейн М. И. Переходные процессы в гребенчатых фильтрах. «Радиотехника», 1957, № 7.

57. Фурдуев В. В. О некоторых понятиях теории сигналов. «Радиотехника», 1957, № 4.

58. Зиберт. Общие закономерности обнаружения целей при помощи радиолокации. «Вопросы радиолокационной техники», 1957, № 5.

59. Харкевич А. А. Об одной схеме приема сигналов. «Электросвязь», 1957, № 2.

60. Харкевич А. А. О теоретически оптимальной системе связи. «Электросвязь», 1957, № 5.

61. Харкевич А. А. О возможностях сжатия спектра сигнала. «Электросвязь», 1957, № 4.

62. Фалькович С. Е. О точности отсчета координаты дальности в радиолокационных системах при некогерентном накоплении. «Радиотехника и электроника», 1957, № 5.

63. Флейшман Б. С. Об оптимальном детекторе с log I_0-характеристикой для обнаружения слабого сигнала при наличии шума. «Радиотехника и электроника», 1957, № 6.

64. Леонов Ю. П. и Телькснис Л. П. Оценка параметров закона распределения случайной функции при ограниченных априорных данных. «Автоматика и телемеханика», 1957, № 11.

65. Харкевич А. А. О теореме Котельникова. «Радиотехника», 1958, № 8.

66. Турбович И. Т. К вопросу о применении теоремы Котельникова к функциям времени с ограниченным спектром. «Радиотехника», 1958, № 8.

67. Вощицкий В. С., Слуцковский А. И. Схема для измерения слабых сигналов со сплошным спектром. «Радиотехника», 1958, № 9.

68. Мешковский К. А. Вопросы помехоустойчивости систем связи, осуществляющих прием сигнала в «целом». «Радиотехника», 1958, № 6.

69. Поляк Ю. В. и Кельзон В. С. К теории обнаружения периодических импульсных сигналов в гауссовом шуме при некогерентном накоплении. «Радиотехника и электроника», 1958, № 6.

70. Тарасенко Ф. П. Об информативности параметров принимаемого сигнала. «Радиотехника и электроника», 1958, № 4.

71. Лезин Ю. С. О синтезе фильтров, оптимальных импульсам определенной формы. Научные доклады Высшей школы, серия «Радиотехника и электроника», 1958, № 3.

72. Сифоров В. И. О пропускной способности каналов связи с медленными случайными изменениями параметров. Научные доклады Высшей школы, серия «Радиотехника и электроника», 1958, № 3.

73. Филиппов Л. И. Идеальное радиоприемное устройство выявления сигналов. Научные доклады Высшей школы, серия «Радиотехника и электроника», 1958, № 2.

74. Железнов Н. А. Принцип дискретизации стохастических сигналов с неограниченным спектром. «Радиотехника и электроника», 1958, № 1.

75. Зубков В. Д. Оптимальное обнаружение при коррелированных помехах. «Радиотехника и электроника», 1958, № 12.

76. Школьник. Обнаружение импульсных сигналов в шумах. «Вопросы радиолокационной техники», 1958, № 3.

77. *Миллер, Бернштейн.* Теория когерентного интегрирования и ее применение к обнаружению сигналов. «Вопросы радиолокационной техники», 1958, № 5.

78. Бласбалг. Последовательное обнаружение в гауссовом шуме. «Вопросы радиолокационной техники», 1958, № 4.

79. Добрушин Р. Л. Одна статистическая задача теории обнаружения сигнала на фоне шума в многоканальной системе. «Теория вероятности и ее применения», 1958, № 2. /443

80. Харкевич А. А. Опознание образов. «Радиотехника», 1959, № 5.

81. Финкельштейн М. И. Оптимальные полосы пропускания частот гребенчатых фильтров. «Радиотехника», 1959, № 1.

82. Фалькович С. Е. Некоторые результаты применения метода апостериорной вероятности к задачам проектирования радиолокационных систем. «Радиотехника и электроника», 1959, № 4.

83. Зубаков В. Д. Обнаружение сигналов на фоне нормальных шумов и хаотических отражений. «Радиотехника и электроника», 1959, № 1.

84. Зубаков В. Д. Обнаружение когерентных сигналов на фоне коррелированных помех. «Радиотехника и электроника», 1959, № 4.

85. Башаринов А. Е. и Флейшман Б. С. Применение метода последовательного анализа в системах двухзначной передачи при релеевских флуктуациях интенсивности сигналов. «Радиотехника и электроника», 1959, № 2.

86. Сраговнч В. Г. Об оптимальном обнаружении сигнала на фоне коррелированной гауссовой помехи. «Радиотехника и электроника», 1959, № 5.

87. Вайнштейн Л. А. Радиолокационное обнаружение «мерцающего объекта» на фоне коррелированных помех. «Радиотехника и электроника», 1959, № 5 и № 7.

88. Башаринов А. Е., Флейшман Б. С., Самохина М. А. Бинарные накопительные системы с двухпороговыми анализаторами. «Радиотехника и электроника», 1959, № 9.

89. Гуткин Л. С. О значении критериев оптимальности и априорных распределений в теории приема сигналов. «Радиотехника и электроника», 1959, № 10.

90. Финк Л. М. О потенциальной помехоустойчивости при замираниях сигнала. «Радиотехника», 1959, № 9.

91. Финк Л. М. О потенциальной помехоустойчивости при неопределенной фазе сигнала. «Радиотехника», 1959, № 1.

92. Тарасенко Ф. П. Сравнение методов радиолокационного приема с точки зрения теории информации. «Радиотехника», 1959, № 7.

93. Варшавер Б. А. К теории передачи сигналов со многими дискретными значениями. «Радиотехника», 1959, № 1.

94. Клячкин Л. З. Помехоустойчивость автокорреляционного приемника АМ сигналов. «Радиотехника», 1959, № 2.

95. Турбович И. Т. Аналитическое представление функций времени с неограниченным спектром. «Радиотехника», 1959, № 3.

96. Хургин Я. И. Оценка пропускной способности некоторых каналов связи со случайно изменяющимися параметрами. «Радиотехника», 1959, № 12.

97. Аптэк Ю. Э. Гребенчатый фильтр с электронным коммутатором. «Радиотехника и электроника», 1959, № 11.

98. Добрушин Р. Л. Оптимальная передача информации по каналу с неизвестными параметрами. «Радиотехника и электроника», 1959, № 12.

99. Ширман Я. Д. Теория обнаружения полезного сигнала на фоне гауссовых шумов и произвольного числа мешающих сигналов со случайными амплитудами и начальными фазами. «Радиотехника и электроника», 1959, № 12.

100. Черняк Ю. Б. О некоторых способах обработки флуктуирующих сигналов в двухканальных системах. «Радиотехника и электроника», 1959, № 12.

101. Дубинский Б. А. К вопросу о точности измерения параметров колебания, искаженного гауссовой помехой малой интенсивности. «Радиотехника и электроника», 1959, № 12.

102. Уолтер. Количественный анализ автоматических схем обнаружения и оценки положения мерцающих целей при наличии шума. «Радиотехника и электроника за рубежом», 1959, № 1.

103. Зиберт. Некоторые применения теории обнаружения к радиолокации. «Радиотехника и электроника за рубежом», 1959, № 1.

104. Каширин В. А., Шастова Г. А. Возможности повышения помехоустойчивости на основе использования распределения вероятностей параметра. «Автоматика и телемеханика», 1959, № 9.

105. Воюцкий В. С. О сравнительной помехоустойчивости двухканального корреляционного приемника с квадратичным детектором. «Радиотехника», 1960, № 1.

106. Гуткин Л. С. Некоторые соотношения в оптимальных системах обнаружения сигналов. «Радиотехника», 1960, № 2 и № 4.

107. Черняк Ю. Б. Приближенный метод расчета характеристик обнаружения многоканальных систем с коррелированными шумами при отборе амплитуд по наибольшему значению. «Радиотехника и электроника», 1960, № 2.

108. Митяшев Б. Н. Об оптимальном способе определения временного положения импульса. «Радиотехника и электроника», 1960, № 2.

109. Бакаринов А. Е. Обнаружение импульсных пакетов случайной продолжительности устройствами с конечной емкостью памяти. «Радиотехника и электроника», 1960, № 3.

110. Черняк Ю. Б. Обнаружение сигнала с неизвестной частотой и произвольной начальной фазой на фоне шума. «Радиотехника и электроника», 1960, № 3.

111. Мельников В. С. Разнос сигналов частотного телеграфирования при замираниях и идеальном приеме. «Электросвязь», 1960, № 3.

112. Гуткин Л. С. Зависимость параметров детектирования и преобразования частоты от амплитуд сигнала и гетеродина. «Радиотехника», 1946, № 6.

113. North D. O. Analysis of factors which determine signal-to-noise discrimitation in pulsed carrier systems, Rep. PTR-6C, RCA, Princeton, 1943.

114, Dwork B. M. Proc. IRE, 1950, v. 38, p. 771.

115. Гренадер У. Случайные процессы и статистические выводы. Изд-во иностранной литературы, 1961.

116. Zaden L., Ragazzini J. An extension of Wieners theory of prediction. J. of Appl. Phys., 1950, v. 21, № 7.

117. Booton R. An optimization theory for time-varying linear systems with non-stationary statistical inputs. Proc. IRE, 1959, № 6.

118. Миддлтон Д. и Ван-Метер Д. Обнаружение и воспроизведение сигналов, принятых на фоне шумов, с точки зрения теории статистических решений. В сборнике переводов «Прием импульсных сигналов в присутствии шумов». Госэнергоиздат, 1960.

119. Slepian D. Estimation of signal parameters in the presence of noise. Trans. IRE, PGIT, 1954, № 3.

120. Middleton D. A note of the estimation of signal waveform. Trans IRE, IT, 1959, № 2.

121. Blasbalg H. Experimental results in sequential detection. Trans. IRE, IT, 1959, № 2.

122. Шапиро И. И. Расчет траекторий баллистических снарядов по данным радиолокационных наблюдений. Изд-во иностранной литературы, 1961.

123. Сраговић В. Г. О расчете характеристик обнаружения при квадратичном суммировании сигналов. «Радиотехника и электроника», 1960, № 4.

124. Левин Б. Р. Оптимальные фазовые методы обнаружения сигналов. «Радиотехника и электроника», 1960, № 4.

125. Гуткин Л. С. Теория оптимальных методов радиоприема при флуктуационных помехах. Госэнергоиздат, 1961.

126. Бакаринов А. Е., Флейшман Б. С. Методы статистического последовательного анализа и их приложения. Изд-во «Советское радио», 1962.

127. Бакут П. А. и др. Вопросы статистической теории радиолокации. Т. I. и II. Изд-во «Советское радио», 1962, 1963.

128. Хелстром К. Статистическая теория обнаружения сигналов. Изд-во иностранной литературы, 1963.

129. Ширман Я. Д., Голиков В. Н. Основы теории обнаружения радиолокационных сигналов и измерения их параметров. Изд-во «Советское радио», 1963.

130. Тихонов В. И. Статистическая радиотехника. Изд-во «Советское радио», 1966.

131. Левин Б. Р. Теоретические основы статистической радиотехники. Книга вторая. Изд-во «Советское радио», 1968.

132. Цыпкин Я. З. Адаптация и обучение в автоматических системах. Изд-во «Советское радио», 1968.

133. Большаков И. А., Гуткин Л. С., Левин Б. Р., Стратонович Р. Л. Математические основы современной радиоэлектроники. Изд-во «Советское радио», 1968.

134. Фалькович С. Е. Оценка параметров сигнала. Изд-во «Советское радио», 1970.

135. Вакман Д. Е. Сложные сигналы и принцип неопределенности в радиолокации. Изд-во «Советское радио», 1965.

136. Левин Ю. С. Оптимальные фильтры и накопители импульсных сигналов. Изд-во «Советское радио», 1963.

137. Миддлтон Д. Очерки теории связи. Изд-во «Советское радио», 1966.

138. Стратонович Р. Л. Условные марковские процессы и их применение к теории оптимального управления. Изд. МГУ, 1966.

139. Амиантов И. Н. Применение теории решений к задачам обнаружения сигналов и выделения сигналов из шумов. Изд. ВВИА им. Жуковского, 1958.

140. Вентцель Е. С. Введение в исследование операций. Изд-во «Советское радио», 1964.

141. Мак-Кинси Дж. Введение в теорию игр. Физматгиз, 1960.

142. Льюс Р., Райфа Х. Игры и решения. Изд-во иностранной литературы, 1961.

143. «Бесконечные антагонистические игры». Сборник статей под ред. Н. Н. Воробьева. Физматгиз, 1963.

144. Гуткин Л. С., Борисов Ю. Г и др. Радиоуправление реактивными снарядами и космическими аппарата... Изд-во «Советское радио», 1968.

145. Блекуэлл Д., Гиршик М. А. Теория игр и статистических решений. Изд-во иностранной литературы. 1958.

146. Себестиан Г. С. Процессы принятия решений при распознавании образов. Изд-во «Техника», Киев, 1965.

147. Барабаш Ю. Л. и др. Вопросы статистической теории распознавания. Изд-во «Советское радио», 1967.

148. Линник Ю. В. Метод наименьших квадратов и основы теории обработки наблюдений. Физматгиз, 1962.

149. Беллман Р. Динамическое программирование. Изд-во иностранной литературы, 1960.

150. Линник Ю. В. Статистические задачи с мешающими параметрами. Изд-во «Наука», 1966.

151. Леман Э. Проверка статистических гипотез. Изд-во «Наука», 1964.

152. Ван-дер-Варден Б. Л. Математическая статистика. Изд-во иностранной литературы, 1960.

153. Ли Р. Оптимальные оценки; определение характеристик и управление. Изд-во «Наука», 1966.

154. Стратонович Р. Л. Применение теории процессов Маркова для оптимальной фильтрации сигналов. «Радиотехника и электроника», 1960, № 11.

155. Стратонович Р. Л. Теория оптимальной нелинейной фильтрации случайных функций. «Теория вероятностей и ее применения», 1959, № 2.

156. Большаков И. А., Репин В. Г. Вопросы нелинейной фильтрации. «Автоматика и телемеханика», 1961, № 4.

157. Гуткин Л. С. Потенциальная точность измерения в одноканальных и многоканальных измерителях параметров сигнала. «Радиотехника», 1964, № 3 и 4.

/446

158. Стратонович Р. Л. Оптимальный прием узкополосного сигнала с неизвестной частотой на фоне шумов. «Радиотехника и электроника», 1961, № 7.

159. Гуткин Л. С. Сравнение реальной и потенциальной точностей пеленгации. «Радиотехника», 1965, № 6 и 1966, № 3.

160. Стратонович Р. Л. Существует ли теория синтеза оптимальных адаптивных, самообучающихся или самонастраивающихся систем? «Автоматика и телемеханика», 1968, № 1.

161. Цыпкин Я. З. А все же существует ли теория синтеза оптимальных адаптивных систем? «Автоматика и телемеханика», 1968, № 1.

162. Репин В. Г., Тартаковский Г. П. Адаптация систем приема и обработки информации и теория статистических решений. «Автоматика и телемеханика», 1968, № 3.

163. Корадо В. А. Оптимальное обнаружение случайных сигналов на фоне случайных помех неизвестной интенсивности при условии постоянства вероятности ложной тревоги. «Радиотехника и электроника», 1968, № 5.

164. Корадо В. А. Оптимальное обнаружение детерминированных сигналов со случайными параметрами на фоне помех с неизвестной интенсивностью при постоянной вероятности ложной тревоги. «Радиотехника и электроника», 1968, № 6.

165. Дмитриенко А. Н., Корадо В. А. Характеристики обнаружения когерентной пачки импульсов с известной начальной фазой на фоне гауссовой помехи с неизвестной интенсивностью. «Радиотехника и электроника», 1968, № 9. Там же статья тех же авторов об обнаружении пакета независимо флуктуирующих импульсов.

166. Дмитриенко А. Н., Корадо В. А. Характеристика обнаружения когерентной пачки импульсов со случайной начальной фазой на фоне гауссовой помехи с неизвестной интенсивностью. «Радиотехника и электроника», 1968, № 12.

167. Корадо В. А. Об оптимальном обнаружении сигналов при действии помех с неизвестными параметрами. «Радиотехника и электроника», 1969, № 2.

168. Левин Б. Р., Кушнир А. Ф. Асимптотически оптимальные алгоритмы обнаружения и различения сигналов на фоне помех. «Радиотехника и электроника», 1969, № 2.

169. Левин Б. Р., Кушнир А. Ф. Асимптотически оптимальные ранговые алгоритмы обнаружения сигналов на фоне помех. «Радиотехника и электроника», 1969, № 2.

170. Левин Б. Р., Троицкий Е. В. Полная вероятность ошибки при классификации нормальных совокупностей, различающихся векторами средних. «Автоматика и телемеханика», 1970, № 1.

171. Гаджиев М. Ю. Применение теории игр к некоторым задачам автоуправления. «Автоматика и телемеханика», 1962, № 8 и 9.

172. Цыпкин Я. З. Адаптация, обучение и самообучение в автоматических системах. «Автоматика и телемеханика», 1966, № 1.

173. Гуткин Л. С. Теория оптимальных методов радиоприема (раздел обзора развития теории информации в СССР). «Техническая кибернетика», 1963, № 5.

174. Гуткин Л. С. О законе установления дисперсии шума на выходе согласованного линейного фильтра. «Известия вузов». Радиоэлектроника, 1970, № 2.

175. Финк Л. М. Теория передачи дискретных сообщений. Изд-во «Советское радио», 1970.

176. Вольмаков И. А. Выделение потока сигналов из шума. Изд-во «Советское радио», 1969.

177. Пестряков В. Б. Фазовые радиотехнические системы. Изд-во «Советское радио», 1968.

178. Финкельштейн М. П. Гребенчатые фильтры. Изд-во «Советское радио», 1969.

179. Кузьмин С. З. Цифровая обработка радиолокационной информации. Изд-во «Советское радио», 1967. /447

180. Самсоненко С. В. Цифровые методы оптимальной обработки радиолокационных сигналов. Воениздат, 1968.

181. Черняк Ю. Б. О линейных свойствах системы широкополосный ограничитель—фильтр. «Радиотехника и электроника», 1962, № 7.

182. Бородин Л. Ф. Введение в теорию помехоустойчивого кодирования. Изд-во «Советское радио», 1968.

183. Шастова Г. А. Кодирование и помехоустойчивость телемеханической информации. Изд-во «Энергия», 1966.

184. Стратонович Р. Л. О ценности информации. «Техническая кибернетика», 1965, т. 5.

185. Шеннон К. Работы по теории информации и кибернетике. Изд-во иностранной литературы, 1963.

186. Варшавер В. А. К сравнению способов приема сигналов. «Радиотехника», 1962. № 2.

187. Шаров А. И. Идеальный прием сигналов оптимального кода «Радиотехника», 1961, № 9.

188. Пестряков В. Б. Оптимальное обнаружение радиосигналов, ч. I. Изд. МЭИС, 1967.

189. Петрович Н. Т., Размахнин М. К. Системы связи с шумоподобными сигналами. Изд-во «Советское радио», 1969.

190. Фомин А. Ф. Пороговые свойства и оптимальные параметры сигналов при радиолокационных измерениях. «Радиотехника», 1968, № 6.

191. Фомин А. Ф. Оценка достоверности передачи сообщений при использовании аналоговых широкополосных сигналов. «Радиотехника», 1970, № 5.

192. Robbins H., Monro S. A stochastic approximation method. Annales of Math. Statistics, 1951, v. 22, № 1.

193. Kiefer E., Wolfowitz J. Stochastic estimation of the maximum of regression functions. Annales of Math. Statistics, 1952, v. 23, № 3.

194. Питерсон И. Л. Статистический анализ и оптимизация систем автоматического управления. Изд-во «Советское радио», 1964.

195. Логинов Н. В. Методы стохастической аппроксимации. «Автоматика и телемеханика», 1966, № 4.

196. Корадо В. А. Об оптимальном обнаружении сигналов на фоне помех с неизвестными параметрами при ограниченной вероятности ложной тревоги. «Радиотехника и электроника», 1970, № 7.

197. Большаков И. А. и др. Некоторые вопросы статистического синтеза информационных систем. «Техническая кибернетика», 1970, № 2.

198. Цыпкин Я. З. Основы теории обучающихся систем. Изд-во «Наука», 1970.